Loughborough
University

This item was submitted to Loughborough's Institutional Repository by the
author and is made available under the following Creative Commons Licence
conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

Peer assessment or promotion by numbers?

A comparative study of different measures of researcher performance within the UK Library and Information Science research community

by

David Edward Bennett

An edited version of a Master's Dissertation.
The original Dissertation was submitted in partial
fulfilment of the requirements for the award of
Master of Science degree of Loughborough
University

September 2007

Supervisor   Professor Charles Oppenheim
Department of Information Science

Abstract

Hirsch's *h*-index, Egghe's *g*-index, total citation and publication counts, and five proposed new metrics were correlated with one another using Spearman's Rank Correlation for one hundred randomly selected academics and researchers working in UK Library and Information Science departments. Metrics were compared for individuals of different genders and at institutions awarded different RAE (2001) grades. Individuals' metrics were rank-correlated against academic ranks and RAE (2001) grades of their employing departments. Metrics calculated using *Web of Science* and *Google Scholar* data were compared. Peer- and *h*-index metric-ranked orders of researchers were rank-correlated. Citation behaviour and attitudes towards peer and citation-based assessment of 263 academics and researchers were investigated by factor analysis of online attitudinal survey responses.

h increased curvilinearly with total citation and publication counts, suggesting that *h* was constrained by the activity in the field preventing individuals producing enough heavily cited publications to increase their *h*-index scores. Most individuals therefore shared similar *h*-index scores, making interpersonal comparisons difficult. Total citation counts and Bihui's *a*-index scores distinguished between more individuals, though whether they could confidently identify differences between individuals is uncertain. Both databases arbitrarily omitted individuals and publications, systematically biasing citation metrics calculated using them.

In contrast to studies of larger fields, no citation metrics correlated with RAE grade, academic rank, or direct peer-assessment, suggesting that citation-based assessment is unsuitable for research fields with relatively little research activity. No gender bias was evident in academic rank, esteem or citedness.

At least nine independent factors influence citation behaviour. Mertonian factors dominated. The independence of the factors suggested different individuals have different combinations of non-Mertonian motivations. The overriding meaning of citations was confirmed as signals of relevance and reward.

Recommendations for future research include a need to develop simple, robust methods to identify subfields and normalise citations across subfields, to quantify the impact of random bias and to determine whether it varies across subfields, and to study the rate of accumulation of citations and citation distribution changes for individuals (and departments) over time to determine whether career age can be controlled for, in particular.

Acknowledgements

The unwavering assistance, patience and support of Professor Charles Oppenheim throughout this dissertation is gratefully acknowledged. Grateful thanks are extended to all those who responded to the questionnaire.

Contents

List of figures

xiv

List of tables

1   Introduction

1.1   Background

1.1.1 Political interest in quantitative research assessment

Funding bodies need to demonstrate that they spend public money in the most cost-effective manner. Researchers therefore have to convince sponsors lacking specialised knowledge of their field of work that their proposed research project will have a significant social or technological payoff (Kostoff 1998, p. 29).

Increasing demands to objectively justify publicly funded research (Kelly & Jennions 2006, p. 167; García-Aracil, Gracia & Pérez-Marín 2006, p. 213), expressed in the UK through the Research Assessment Exercise (RAE), interested funding bodies in the possibility of cheaply and quickly calculated metrics (Saad 2006, p.119; Norris & Oppenheim 2003, p. 710) that may reliably estimate the research impact of different individuals (Oppenheim 2006, p. 4), research groups (van Raan 2006, p. 496) and institutions (Oppenheim 1997, p. 478). Citation metrics have been particularly well accepted by officials and fundholders who now plan to largely replace peer assessment with comparison of citation metrics (*Science and innovation investment framework 2004-2014: next steps* 2006, p. 31) because they are cheap (Oppenheim 1997, p. 485) simple, heuristic, numeric measures with apparent validity.

Following the failure of publication counts as a measure of research performance (Hargens & Schuman 1990, p. 205), citation counts were investigated as impact metrics. Traditional interpretations of citation counts based upon Merton's construct that citations form a rhetorical reward system have been attacked on the basis that publications never cite all relevant publications and because little empirical research has been conducted to investigate which factors affect the

likelihood of a publication being located, or on the mechanics of citation, *i.e.* what characteristics of publications influence relevance judgements, recall, and choice of which publications to cite (*e.g.* Bornmann & Daniel 2007, p. 1384; Zuckerman 1987, p. 336).

1.1.2 Novel citation metrics

Hirsch (2005) proposed a new citation metric, the *h*-index, defined as being the number of publications in a collection (usually the lifetime output of an individual or research body) that have each been cited at least *h* times. Whilst it has been established as a valuable heuristic measure, like total citation counts, *h* has been attacked repeatedly for over-simplifying research performance (Bornmann & Daniel 2007, p. 1384). Whilst various modifications of the *h*-index have been attempted to normalise *h* for different numbers of authors (Batista *et al.* 2005, p. 179) and to weight the *h*-index to better represent heavily cited publications (Bihui 2006 *cited in* Bornmann & Daniel 2007, p. 1384), no attempt has yet been made to extract the different facets represented by *h* to form a meaningful set of *h*-based metrics that are more informative than other metrics (van Raan 2006, p. 501). A series of *h*-derived metrics were therefore constructed to extract and explore the behaviour of different facets of the *h*-index.

Since citation metrics are proposed as a means to quantify research performance, it would be useful to compare metrics against local peer assessment, academic rank and Research Assessment Exercise grades to establish whether they might measure the same thing.

1.1.3 Database coverage bias

Google Scholar is a serious competitor to *Web of Science* as a data source for citation analysis, both because of concern over the selective coverage of ISI (MacRoberts & MacRoberts 1996, p. 438) compared to *Scopus* or *Google Scholar*

(Yang & Meho 2006, p. 10), and because open access institutional repositories are being designed to be hospitable to *Google* (Getz 2005, p. 31). If the serials crisis forces libraries to switch from traditional subscription models to a "golden" road open access model (Harnad *et al.* 2004, p. 314), reliable citation metrics will be necessary to assist administrators.

1.2 Aim

To investigate citation behaviour and citation metrics, and to compare citation and peer-based assessment in Library and Information Science in the UK.

1.3 Objectives

- To explore the validity and reliability of citations, and whether they may act as valid indicators of research performance

- To investigate the characteristics of proposed citation metrics as measures of researcher performance

- To assess the perceived value of citation metrics and peer assessment amongst academic Library and Information Management researchers

- To explore which factors determine whether articles are cited and whether factors are predictably associated with one another

- Propose and test several new citation metrics (Appendix A):

$$v = \frac{h}{p_{\mathrm{tot}}}$$
where h = Hirsch's index
p_{tot} = total number of publications

$$\Lambda = \frac{g}{h}$$
where g = Egghe's g
h = *Hirsch's h*

$$h^{+} = \sum_{i=1}^{h} c_i - h^2$$
where the publications are arranged in decreasing order of citedness
c_i = number of citations for each individual publication
h = Hirsch's h

$$s = \frac{1}{c_{tot}} \sum_{i=1}^{h} c_i \qquad \text{where} \quad \text{the publications are arranged in decreasing order of citedness}$$

c_i = number of citations for each individual publication

c_{tot} = total number of citations of all publications

h-range = difference between maximum and minimum number of citations contributing to h

M_c = median number of citations of all cited publications

M_o = median number of citations of all publications

- To calculate and compare h, g, a, total citation counts, total publication counts, median citation counts, mean numbers of citations per publication, median numbers of citations per publication, v, s, Λ, h^+, h-range, M_c and M_o indices for a randomly selected sample of UK Library and Information Science academics

- To compare the reliability of the Thomson ISI *Web of Science* Science, Social Sciences and Arts and Humanities citation indices with *Google Scholar*, as accessed through the *ScHolar Index* web platform

- To survey senior UK-based Library and Information Management researchers using an online questionnaire comprising multiple cryptic questions answered using Likert scales, to investigate the factors influencing why researchers cite articles, and whether prestige of publication or parity of esteem with authors affects the likelihood that an article will be cited using factor analysis

- To identify concerns amongst UK Library and Information Science researchers about peer assessment and citation analysis as means of research performance review

- To determine whether citation metrics vary with academic rank

- To determine whether there is a gender bias in citation, promotion or prestige amongst UK Library and Information Science academics

- To identify any association between the average citation metric of employees in Library and Information Science Departments and the Research Assessment Exercise grade of those departments

- To determine whether h based rankings correlate with peer rankings

2 Literature review

2.1 Introduction

The scientific community has always been socially stratified (Cole & Cole 1973). Traditionally, stratification – demonstrated by promotion, prestige and tenure – was based upon publication counts. This was a reasonable approach, since publications declare new discoveries, and "making new discoveries is the goal of the scientist". (Cronin 1984, p. 1)

Individual research does not happen in a vacuum, however (Cronin 1984, p. 1). In addition to making discoveries and publishing their results, researchers must exert their moral rights to be recognised as the person who made the discovery. Merton proved that science builds on past discoveries, each researcher giving up their intellectual property rights in order to publish, in exchange for their ideas to be noticed and discussed within the scholarly community (Cole & Cole 1973, p. 58) through each publication citing all the publications to which it owes an "intellectual debt" (Cronin 1984, p. 14; Ahmed *et al*. 2004, p. 148; White & Wang 1997, p. 130-132). The number of citations pointing towards works by any one author, research group or institution demonstrates their relative importance to a particular field or subfield of research.

Since no author could reasonably be expected to reference all the publications that influenced a work, references in publications represent selective "votes" of confidence, signifying trust and thanks to authors of publications considered particularly important to the citing work (Cronin 2005, p. 143; Kurtz *et al.* 2004 *cited in* Cronin 2005, p. 72). This is a strategy where prolonged, reciprocal relationships ensure that all citations indicate some intellectual debt.

Psychological factors also determine which publications are cited (Harter, Nisonger & Wang 1993 *cited in* Cronin 2005, pp. 153-154). Brooks (1986, p. 35) found that five of the seven factors that predicted citation in Library and Information Science in the USA, currency, persuasiveness, reward, awareness raising and indicating consensus, were all strongly correlated with each other, indicating an overarching factor of "persuasiveness", a view supported more generally by Cozzens (1988, p. 441). Persuasiveness was negatively correlated with the other two factors identified, criticism and functional citation, suggesting that persuasiveness was associated with seeking scholarly consensus, a view confirmed by White & Wang (1997, p. 130-132). Relevance judgements also influence citation behaviour (Ghaebi 2003, p. 112; Harter 1992, p. 603).

Citations both indicate the impact of a research publication and advertise its existence, influencing the visibility of a researcher's work, which is the main reward in research (Cole & Cole 1973, p. 58), and which predicts prestige and position better even than publication counts (Cole & Cole 1973, p. 27). Recently, several new citation metrics have been suggested to evaluate research performance.

The reliability and assumptions of citation analysis as a research assessment tool have been attacked by MacRoberts & MacRoberts (1986), and more recently by Döring (2007), on the grounds that the reasons for citing publications are inconsistent, casting doubt on the validity and reliability of citation analysis as a method of research assessment.

White (2004, p. 115) found that the Mertonian pattern of citations as acknowledgements of intellectual debt fitted that found in the literature better than models that assume individuals all attempt to cite influential figures and works in order to attract more attention and citations themselves. Since misleading work by fraudsters is excised from the literature after their exposure (*e.g.* Kochran & Budd 1992, p. 491), this suggests that prolonged, reciprocal relationships effected through peer review ensures that all citations indicate some intellectual debt. This creates an

audit trail of accountability for ideas (Cronin 2005, p. 15) and expertly indexes the literature (Brooks 1986, p. 34). Slight deviations are possible from the Mertonian norms but over large groups over a period of time, citations are considered to be reliable indicators (van Raan 2000, p. 306).

2.2   Suggested violations of Mertonian norms

Nederhof and van Raan (1987*a*, p. 326) suggested that both systematic and random biases affect the citation counts of individuals and groups. Random factors should produce no systematic bias against any one research group over a sufficiently long period of time (Kostoff 1998, p. 33; Cole & Cole 1977, p. 32), but they might be larger than the difference between individuals' citation metrics, particularly amongst researchers of younger career age (Garfield & Welljams-Dorof 1992, p. 323) whose citation counts are smaller relative to the size of the possible errors. Van Raan (2005*a*, p. 140) advised that authorised publication lists be used to eliminate homonyms (works by different researchers who share the same name) for assessments.

2.3   Random biases

2.3.1 Random miscitations

Citations are often misattributed (MacRoberts & MacRoberts 1986, p. 158), often as a result of the inclusion of homonyms, changes of professional name and human error, whilst 7-20% of citations in databases contained typographical errors, although this is expected to have decreased after electronic updating was introduced (Baird & Oppenheim 1994, p. 6; Meho 2007, p. 32). Provided authorised publication lists are used to eliminate homonyms and systematic bias against individuals with unusual name spellings, as suggested by van Raan (2005*a*, p. 140), the errors are unlikely to significantly bias comparisons across entire research groups.

2.3.2 Secondary citations

Secondary citations are citations to a publication describing the findings of an earlier work (MacRoberts & MacRoberts 1987*a*, p. 305). Whilst citing influential publications may allow individuals to divert attention from the original work to their own (Cozzens 1988, p. 443), there is no evidence that secondary citations systematically favour one author or research group over another, and the secondary citations still accurately indicate intellectual debt to the secondary source. Cronin (*e.g.* Cronin, 2005, p. 143) suggested that acknowledgements are more important recognitions of intellectual input than citations, yet these are never indexed. His views on acknowledgments have not, however, been applied by others.

2.3.3 Negational citations

Cronin (2000, p. 440) and Hart (2007, p. 192) argued that some citations signal praise, whilst others criticise the cited publication, confusing the meaning of citation metrics. Many citations are accompanied by some critical appraisal, whilst very few citations are purely critical (MacRoberts & MacRoberts 1987*b*, p. 456), perhaps because peer review prevents the publication of poor work (Boyack & Börner 2003, p. 447; Baird & Oppenheim 1994, p. 7) and all citations indicate relevance and influence. "Negational" citations are also rare (Oppenheim 1996, p. 158; MacRoberts & MacRoberts 1987*a*, p. 456) because poor papers are generally ignored rather than criticised.

2.4 Systematically acting biases

2.4.1 Geographical publication bias

Articles are predominantly cited in the same country as they were originally published. For example, 31.6% more citations in cardiovascular medicine originated from the same country as the citing article (mainly the USA and UK) than could be expected by chance (West & McIlwaine 2002, p. 503; Pasterkamp *et al.* 2007, p.

163). Pasterkamp *et al.* (2007, p. 154) suggested that the differences in the levels of citation between countries could be explained by differences in national publication volume bias and the visibility of publications through personal communications at national meetings. Wealthier countries can also afford to fund expensive world class research, host a greater proportion of all scholarly journals (Trimble 2005, p. 414) and meetings, allowing the selective dissemination of research amongst compatriots, affording English speaking countries a competitive advantage in disseminating research in a timely manner (Liu 1993*a*, p. 21; Kim 2004, p. 91; Pasterkamp *et al.* 2007, p. 163).

2.4.2 Gender bias

Both methods of assessment appear susceptible to sexual discrimination (Motluk 1997 *cited in* Garfield 1998, p. 78; Ferber 1986, p. 382), favouritism (Garfield 1998, p. 78) and influence, but these effects should be less pronounced in citation metrics because with the exceptions of research groups that publish all the work in a particular subfield, citations integrate the opinion of many different researchers. Ferber (1986, p. 388) argued that the apparent citation discrimination she observed may have resulted from researchers associating and disseminating their research more with researchers of their own gender.

2.4.3 Citation database coverage bias

This geographic and English language visibility/journal citation bias is exacerbated by similar coverage biases in the ISI citation indexes (Liang 2006, p. 540; Meho 2007, p. 32). ISI also ignores editorials, claiming them to be "uncitable", together with most proceedings, all books (Manafy 2007, p. 10) and journals it does not consider "core" to a field (MacRoberts & MacRoberts 1996, p. 438) but does now include letters, corrections and retractions (Garfield & Welljams-Dorof 1992, p. 322). This coverage bias may bias citation counts against researchers who publish anywhere other than the "core" journals in their field or in languages other than English. Alternatives to ISI, such as *CiteSeer* and *Scopus* index books, dissertations

and theses (Bar-Ilan 2006, p. 1557) but other than the more extensive coverage of Elsevier journals in *Scopus* (de Moya-Anegón *et al.* 2007, p. 70), the biases of *CiteSeer* are little known.

The coverage of *Scopus* has just been favourably compared to the complete journal listings Ulrich's core journals list. *Scopus* provides a relatively balanced coverage of 120 of Ulrich's 151 areas of research, but its coverage favours sciences over the social sciences and humanitites (de Moya-Anegón *et al.* 2007, p. 75). It includes a similar proportion of publications to Ulrich's for 76 of the 79 countries that it covers, although *Scopus* was found to exhibit a language bias, including a greater proportion of publications written in English, Slovak, Czech and Croation (de Moya-Anegón *et al.* 2007, pp. 60-61). Unlike *Web of Science*, *Scopus* under-represents UK publications (de Moya-Anegón *et al.* 2007, p. 65). Whilst *Scopus* has a less restricted, and therefore less biased, coverage than *Web of Science* (de Moya-Anegón *et al.* 2007, p. 54), it still has significant language and subject bias. Its coverage and biases are different from *Web of Science* as described by Braun, Glänzel & Schubert (2000), however, so comparisons of citation metrics calculated using data from the different databases should be made with caution.

Selective database coverage poses the greatest, and at the institutional level perhaps the only, threat to the reliability of citation metrics based on ISI data. Furthermore the varying coverage of ISI citation indexes over time (Moed 1989, p. 26) means that metrics may not even be comparable over time. For this reason alone, metrics cannot be relied upon as sole indicators of research performance.

2.4.4 Self-citation bias

Self-citation bias, although demonised in the literature (*e.g.* Garfield & Welljams-Dorof 1992, p. 325; Vinkler 2007, p. 481), must be minimal because Saad (2006, p. 118) found that researchers *h*-index scores correlated strongly with raw citation counts (r=0.87, P<0.01), suggesting that the two indices are affected by similar

factors to a similar extent, and because most self-cited works are relatively poorly cited, suggesting that they have a small audience. In many cases, self-cited publications are authored by individuals and research groups that are the only researchers in a subfield and whose own publications are the therefore the only ones available to be cited, for example in psychology (Case & Higgins 2000, p. 640).

2.4.5 Visibility

Where several publications offer the same information, Cozzens (1988, p. 438) argued that the more "convenient", *i.e.* accessible, publication is usually cited. Authors are likely to be exposed to more citations to heavily cited publications and are therefore more likely to remember and cite these publications in their own work, increasing the visibility of the publications yet further. Holden (2006, p. 615) showed that journal impact factors predict the long-term citation rates of articles very accurately, despite the fact that most articles in any journal remain uncited because they are not directly relevant to subsequent research (Seglen 1992, p. 630-631).

Highly cited publications are typically in prestigious, high impact factor journals, and entry into these journals is competitive. Therefore, higher quality articles are published in them. In one study, nine per cent of highly cited doctoral Physics publications appeared in low impact factor journals, suggesting that original high quality research is quickly recognised and cited, increasing its visibility, wherever it is published, although journal impact factor may affect initial rate (Asknes 2003, pp. 163, 168). Of course, the assumption that to get high citation counts one should publish in a high impact factor journal does not follow.

Personal knowledge of the author increases the probability of a publication being cited (Ghaebi 2003, p. 112) and 40% of citations may result from influence relationships (Baird & Oppenheim 1994, p. 6), suggesting that influential authors and "politically correct" works (Kostoff 1998, p. 30) are often cited to increase the

probability of acceptance by high impact factor journals (Daniels *et al.* 2002, p. 266). Nederhof & van Raan (1987*b*, p. 345) showed that this has not resulted in a true "Matthew effect", where individuals of younger career age struggle to become cited.

Highly cited authors appear not to be cited disproportionately often by any given individual (White 2004, p. 108), yet in any field, the bulk of citations are to a small number of very heavily cited authors (White 2004, p. 108; Seglen 1992, p. 636). These authors must therefore be cited infrequently by a large number of individuals.

In principle, peer review constrains influence effects by ensuring that all cited publications are at least partially relevant to the citing work. Even Sokal, who conducted a famous hoax where he published articles including spurious citations to publications outside of the field his referees were familiar with, was constrained in his choice of authors because he had to cite publications that at least appeared relevant to his publications (White 2004, p. 113). Those who have deliberately falsified results, such as Darsee cease to be cited and the literature is purged of their fraudulent work as soon as it is exposed (Kochran & Budd 1992, p. 491), ensuring that cheating is unlikely to succeed. Overall, White (2004, p. 115) found that the Mertonian pattern of citation fitted the literature better than models attempting to maximise persuasiveness. Liu (1993*b*, p. 376) noticed that some publications cited works that were only partially relevant. It is unclear whether the authors always read the works that they cite (Simkin & Roychowdhury [n.d.], p. 1), misread or misunderstood publications (Wright & Armstrong 2007, p. 10), or interpreted partially relevant or specific results as if more general or in the context of the citing work (Leydesdorff & Amsterdamska 1990, p. 319).

Visibility appears to be influenced in part by the impact factor of the journal in which works have been published. This is controlled by peer assessment, which is perhaps influenced by citation and publication counts (Harter & Hooten 1990, p.

13

268). Both peer assessment and citation metrics therefore appear to measure the same parameters and it is unsurprising that their outcomes broadly agree.

2.4.6 Co-authorship increases visibility

There is a consensus that co-authored papers often attract more citations than singly authored papers (Bornmann & Daniel 2006, p. 433; Hart 2007, p. 191; Glänzel & Schubert 2001, p. 213), probably because co-authored publications often utilise shared research expertise to span several research (sub)fields (Seglen 1992, p. 636) and are therefore potentially relevant to, and cited by, more researchers. However, co-authored publications published within the narrowly focussed journal *Academic Library Research* were no more cited than single author works (Hart 2007, p. 193). Inter-disciplinary research would therefore seem to be favoured by citation-based assessment (Döring 2007, p. 709), but even if true this would not necessarily harm scientific enquiry.

2.5   Citation metrics and the *h*-index

Older metrics were criticised by Hirsch (2005, p. 16569) and Kelly & Jennions (2006, p.167) for their bias and susceptibility to manipulation. Hirsch (2005, p. 16569) proposed the *h*-index, *h* being defined as being the largest number of an author's publications that had each been cited at least *h* times. *h* represents the cumulative, broad impact of a researchers' work over time (Hirsch 2005, p. 16569; Liang 2006, p. 153). It is easily calculated from a list of publications ordered by number of citations, a service provided *Web of Science*, *Scopus* and *ScHolar Index* (Manafy 2007).

Hirsch (2005, p. 16569) and Saad (2006, p. 119) claimed that *h*-index scores are more easily checked than other citation metrics because only the citations to publications with just more/less than *h* citations each need be checked to verify the metric but this assumed a skewed citation distribution, such that most of the

publications contributing to h had many more than h citations, otherwise many citations would have to be checked. In fields of low activity, almost all the citations to a group or individual would need to be checked to ensure that miscitations did not affect h. Where they did, the error in h would be larger than for total citation counts (Vanclay 2007, p. 1548), and could be more easily manipulated by strategic self-citation.

h has been criticised for ignoring both highly cited publications (Rousseau [n.d.], p. 4) and the 'long tail' of publications cited fewer than h times, including "premature" papers (Stent 1972 cited in Garfield 1998, p. 75), both of which might be useful in distinguishing between research groups in less active fields. Unlike raw citation counts, the h-index thus risks clustering individuals together into an indistinguishable mass in less active fields, whilst, despite Hirsch's (2005, p. 61569) claims, still arbitrarily and "randomly select[ing] for and against" different researchers of very similar ability depending on whether some works have received just below or above the threshold number of citations to contribute to h. As h-index scores increase over time, so does the threshold number of citations for new publications to contribute to h. The size of the 'long tails' of publications cited fewer than h times are therefore predicted to grow over time, so that an increasing proportion of active researchers' work is ignored by h.

Like all citation metrics, h is retrospective and therefore favours researchers with greater career ages, who have had more time to publish and accrue citations and to develop social networks through which to increase the visibility of their publications.

Hirsch (2005, p. 16571) argued that self-citation could only affect a small number of papers, so the effect would be insignificant, a claim supported by Cronin & Meho (2006) and Oppenheim (2007, p. 298), though whether this holds true for those individuals who cite themselves heavily because they have conducted most of the research in their subfield is unclear.

15

Egghe (2006) proposed the *g*-index, defined as the largest number such that the *g* most cited articles collectively receive at least g^2 citations, as an improvement to the *h*-index that would recognise highly cited works. Like Bihui's *a*-index, calculated as the mean number of citations awarded to an authors *h* most cited papers (Rousseau [n.d.], p. 4; Meho 2007, p. 36), *g* may be biased by only one heavily cited paper, so long as the researcher has published sufficiently many other works, regardless of their quality.

2.6   Citation analysis and peer research assessment

The ranked order of departments following RAE outcomes have been found repeatedly to be and strongly correlated with the ranked total citation counts of departments (*e.g.* Norris & Oppenheim 2003, p. 713; van Raan 2006, p. 496; Seng & Willett 1995, p. 70), suggesting that the proposed shift towards citation-based institutional research assessment (*Science and innovation investment framework 2004-2014: next steps* 2006, p. 31) would not upset the current ranking of departments. This was unsurprising, since citation and publication counts predict the outcome of peer research assessments (Harter & Hooten 1990, p. 268). The strong positive correlation between normalised citation counts, peer ranking and book reviews found by Meho & Sonnenwald (2000, p. 133) suggested that peer assessment draws upon familiarity of individuals' work through recurrent exposure to their publications and references to them. Such familiarity would be largely subfield-dependent and influenced by the same author visibility bias as citation analysis, although since citations take time to accumulate and increase the visibility of a publication, peer assessment may be less retrospective.

Warner (2000, p. 455) warned that the correlations were never perfect and that the gaps between individuals were not the same, and Zhu & Meadows (1991 *cited in* Cronin 2005, p. 127) could not distinguish between two chemistry departments using citation counts, despite them being graded 2 and 5, respectively, in an RAE

assessment, leading Oppenheim (1997, p. 485) to suggest metrics be used to add objectivity and only guide peer review, following the converging partial indicators approach (Cronin 2005, p. 125-126).

However, citations could be considered more valid than peer assessment because they "represent the integrated peer review of everyone in the field" (Wade 1975, p. 430) over time (Bornmann & Daniel 2006, p. 428), whereas peer assessment has been accused of serious subjectivity (Moed 1989, p. 7; Asknes 2006, p. 169).

Even the citation counts of individuals have been associated with peer assessments. Doctoral graduates granted Boehringer Ingelheim Fellowships had significantly higher publication and citation counts than rejected applicants and non-applicants (Bornmann & Daniel 2005a, p. 392) and Asknes (2006, p. 169) found that raw citation counts of publications correlated positively with self-assessments of their contributions to Science (r_S=0.56), despite some researchers' assessments of their work being inversely related to the number of citations received. Wade (1975, p. 430) claimed that a study where two assessments of research articles by peer review correlated with citations analysis more closely than they did with one another was further proof of the validity of citation analysis.

Other studies have suggested that the correlation between peer assessment and citation analysis is weak (r_S=0.2 to 0.4) (Asknes 2006, p. 174), and Pasterkamp *et al*. (2007, p. 162) found that only thirteen of the fifty most respected "landmark" papers published in *JAMA* in 1987 were in the hundred most cited articles in the discipline, perhaps because many important publications were of more interest to practitioners than academics. West & McIlwaine (2002, p. 503) found that the number of citations a paper received in the journal *Addiction* was not correlated with peer evaluation of articles' importance (r_S=-0.02), although the study was small and poorly designed and inter-rater correlation was weak (r_S=0.39) because assessors were required to assess research in subfields with which they were unfamiliar. These last two studies suggest that peer assessment may be subjective and reliable

only when the assessors are familiar with the subfield of the publication being assessed, which might explain the weak correlations in other studies (*e.g.* Zhu & Meadows 1991 *cited in* Cronin 2005, p. 127).

Both raw citation counts and *h*-index scores anticipated (Hirsch 2005, p. 16572) and recognised Nobel Physics prize winners (Cronin 1984, p. 27; Ashton & Oppenheim, 1978), and *h*-index scores were strongly associated with research appraisals for Chemistry researchers applying for research fellowships (Bornmann & Daniel 2007, p. 1382) suggesting that comparison of *h*-index scores (and because they are strongly correlated, citation counts) at any age may be fair, so long as the results are standardised for "scientific age", *e.g.* as in Liang (2006), and further suggesting that citation metrics may serve as indexes of research quality.

Regardless of what citation counts measure, they and RAEs based on informed individual peer assessments produce the same results.

2.7 Do citations risk distortion of scientific enquiry?

Regardless of whether peer assessments and citation metrics agree, the overt reliance on citations performance measures, particularly without subfield normalisation, risks constraining lines of enquiry (van Raan 2005*b*, p. 111) to those that are popular (Cozzens 1988, p. 443), productive, and which are likely to yield positive results that will be easy to publish (Daniels *et al.* 2002, p. 226) and fit easily with other existing publications. Thus, like the current Research Assessment Exercise, they may discourage work in new lines of enquiry, which if unproductive could not be justified whilst existing lines of research were still profitable. In Medicine, the need to publish and be cited already concentrates research and funding on the first promising lines of enquiry even if they are subsequently proved ineffective (Kostoff 1998, p. 34).

Whilst many (*e.g.* Schwartz 1997, p. 24) have argued that uncited research is worthless, uncited publications have been occasionally proven to be of equal quality and more long-term value than other previously heavily cited research (Stent *cited in* Garfield 1998, p. 75).

Assuming that most citations are from publications in the same (sub)field and that the average number of citations per publication remains constant across different subfields, Cole & Cole (1977, p. 32) argued that the number of researchers, citations and publications varies proportionally to the level of research activity in a (sub)field, so that the average number of citations per publication remains constant and that citation counts could therefore be compared across subfields.

Garfield & Welljams-Dorof (1992, p. 322) confirmed that larger, more active fields had higher citation counts, and differences in citation styles across heterogeneous fields mean that publications in different (sub)fields also have different citation rates (Wallin 2005, p. 262; Kostoff 2002, p. 53). Heavily cited publications in fields of different size and activity therefore cannot be directly compared because the maximum number of citations that they could potentially receive varies between (sub)fields.

Non-normalised citations also give less credit to research relevant mainly to professional practice (Schwartz 1997, p. 27; Nederhof & van Raan 1987*a*, p. 327), where research is read and used (Garfield 1997, p. 962; Vinkler 1987, p. 53) but remains uncited due to a lack of later studies.

2.8 Conclusions

Citations therefore comprise a selective reward system that indicates importance and relevance. Citation metrics therefore retrospectively indicate the broad impact of publications on subsequent work. Whilst random biases make comparisons of individuals unreliable, these are believed to cancel one another out across research

groups, making inter-departmental comparisons reliable. Publications that are judged to be of sufficient quality to enter high impact factor journals are more likely to be heavily cited, and heavy citation leads to increased visibility and further citation. Citations thus indicate the relative activity of different lines of scientific enquiry, although not necessarily their long-term worth. Their injudicious interpretation could induce reluctance to investigate new (sub)fields. Variation in activity between subfields requires that citation metrics be normalised before comparisons are made.

3 Methodology

The justification for the chosen methodology is included in Appendix B.

3.1 Sampling

3.1.1 Sampling frame

The sample frame (Appendix C) was all academics, researchers and research assistants in UK university Library and Information Management Departments. Academics in UK Library and Information Schools identified from the CILIP website (http://www.cilip.org.uk/qualificationschartership/Wheretostudy/) were identified from departmental websites and pasted into an Excel 2003 spreadsheet. It was not possible to search the website of every university in the UK to discover if there are academic information science research groups in the UK that are not attached to teaching departments, but this was thought unlikely. Visiting scholars were excluded from the sample frame as many are not permanently resident in the UK. Research students and practitioners interested in research on the Library and Information Research Group (LIRG) JISCmail listserv were excluded because they were considered unrepresentative of academics in general. For convenience, for the remainder of this dissertation the population of academics and research associates will be referred to as "researchers".

Where email addresses for contacts were unavailable or lists comprised academics working in several disciplines with no identification of which were library or information scientists, a more detailed list was requested from an administrator. Where this was refused on the grounds of data protection, the request was either passed on to the academics via the administrator, or academics in the department were emailed and asked for their colleagues' email addresses.

3.1.2 Sampling

For the comparison of metrics, a random sample of one hundred individuals was obtained but the population of 266 was small enough for all researchers to be included in the questionnaire survey.

Appropriate segments within the target population are not obvious and so stratified sampling was not possible. Systematic sampling was avoided because it would have risked introducing a systematic sampling bias if the population (the sampling frame) was recorded in a particular order (Po 1997, p. 4).

Researchers' contact details were recorded in rows in a spreadsheet (Figure 1). A simple random sample was drawn by randomly generating numbers using Microsoft Excel 2003. The researchers corresponding to these random row numbers were then included in the sample.

Since the sample size was large with respect to the underlying population, the risk of random bias resulting from the over-sampling of small population subgroups was minimised (Fowler 2002, p. 30).

Figure 1. Section of sampling frame spreadsheet

3.1.3 Pre-testing

The questions in the survey questionnaire were double-checked to eliminate bias or lack of clarity. The questionnaire was then pre-tested on a convenience sample comprising Professors Oppenheim and Summers of the Department of Information Science, Loughborough University. The pre-test sample was small to minimise response bias if parts of the survey had to be altered as a result of comments. The first question was amended to include an example. Since only small changes were made to the questionnaire following pre-test feedback, the pre-test responses were included in the analysis.

3.2 Questionnaire

3.2.1 Questionnaire design

An online questionnaire was used because it permitted rapid data gathering and anonymity of responses.

The requirements of the online questionnaire software that were deemed necessary were:

- Ability for participants to save their answers and continue later
- Anonymity of responses
- Required questions that had to be answered
- Resilience to use of the web browser "back" button, *i.e.* using it would not abort the survey
- Automated download of all raw data into an Excel spreadsheet
- Control over design, layout, branding, and colour scheme
- Likert scale, tick box and open answers possible

Also desired were:

- Receipt of a confirmatory communication stating the participants who had submitted their forms

- Ability to distribute the questionnaire from one email address and receive replies at another

A suitable application that met these requirements was *Free Online Surveys* (http://freeonlinesurveys.com/). The survey cost £9.95 per month to operate.

3.2.2 Cover email

In order to maximise the number of respondents, all researchers were emailed inviting them to participate in the online survey, a hyperlink to which was included. This cover email (Appendix D) introduced the purpose of the survey and briefly argued why it was important, suggesting an estimated completion times for the questionnaire, and invited queries or comments. A brief summary of the main findings was promised to respondents.

3.2.3 Questionnaire structure

The questionnaire layout was constrained by the software tool employed. Twelve-point, black, "Arial" font was used but the background colour was limited to white with blue dividing lines separating questions. Likert scales were arranged horizontally, beneath the questions. Tick box options were arranged vertically, with answer boxes to the right of their respective answers. Open-ended responses were invited to be entered into white boxes, approximately one-third the width of the screen.

An introduction screen briefly re-iterated anonymity of the survey and outlined the questionnaire structure, and that although the attitudinal questions were all mandatory, the system allowed other questions to be left unanswered. The questions were numbered in the form: "1 of 20" accompanied by a progression indicator, to reduce the risk of participants who had not read the introduction screen becoming frustrated and abandoning the questionnaire.

To avoid context/order effects, all the factor analysis questions (Appendix E) were as specific as possible and used a centrality question format wherever possible, answered using a five-point Likert scale. It was feared that asking researchers to rank their peers might arouse suspicion, so that question was left until almost the end of the survey and the option not to answer it emphasised lest offended participants refused to submit the survey. The attitude question suggesting that peer assessment and citation metrics were being correlated was also positioned at the end because it potentially had implications for the other questions.

The demographic questions were placed at the end, following the advice of Bourque & Fielder (2003, p. 62).

3.2.4 Question design

Every question was designed to be as simple and short as possible. All factor analysis questions were of a centralist attitude type. Every Likert scale contained equal numbers of positive and negative answers. Little additional information was included to minimise the risk of the results being biased by the selective inclusion of arguments.

The open questions were placed at the end to avoid putting respondents off answering the questionnaire.

The invitations to participate were distributed on 20-21 June 2007 and the survey data (Appendix F) was harvested for analysis on 27 August at 5.25 pm.

3.2.5 Correlation analysis

As noted above, a simple random sample of one hundred researchers was selected from the sample frame. It was expected that this method would provide a representative sample, heterogeneous in gender, academic rank, publication volume and citation counts.

MIMAS, who operate Thomson ISI *Web of Science*, confirmed by email[1] that *Web of Science* is updated each Friday. Free text searches for the broad terms "bibliometric*", "neoplasm*", and "red shift" were run daily from Wednesday 19 June. The number of records retrieved increased on Friday, 22 June 2007, confirming that the weekly update had taken place. This was confirmed by ISI (http://portal.isiknowledge.com/portal.cgi).

After the update was confirmed, the *Web of Science* data was collected over five days until Wednesday 27 June, before the next update. *Google Scholar* searches were conducted between 28 June and 2 July. *Google Scholar* was interrogated through the Scholar Index application (http://www-ihm.lri.fr/~roussel/moulinette/ h/h.cgi). All searches were therefore conducted within an eleven day period and were consequently directly comparable.

The publication name of the author was ascertained from the list of publications on their university webpage, if available, or from a *Web of Science* general search of the researcher's name and checked for institution. No homographs were observed in the sampling frame. Even where researchers' publication names were taken from their publication lists, their form varied in both citation indexes, second initials often being omitted. This name was then searched for using the general search in *Web of Science*. An attempt was made to check each individual record for the researcher's work address, to exclude other researchers' works.

---

[1] MIMAS WoK helpdesk support. Email to Prof. C. Oppenheim, 12.07.07.

The "analyse" function was used to select the precise name and then analyse by author searches were used to specify the precise name. The search was then further limited by using the "refine your results" tool to choose to filter results by subject categories. The number of categories displayed was maximised by selecting the "show more" and then "show more categories – up to 100" options. Categories of records were viewed systematically, starting with the categories least likely to contain relevant records. Categories that contained only obvious homographs were excluded. The "citation report" view was then selected to present the remaining records in descending order of the number of times they had been cited. These results were examined and further homographs excluded. No attempt was made to control for self-citation.

When counting the number of authors of papers contributing to a researcher's h-index, where a choice existed between which of several equally cited publications to include in h, the publication with the fewest authors was selected so that the calculated collaboration metric was comparable between researchers. Where a researcher was an editor or compiler of a work, only other editors or compilers were counted.

The total citation counts, total publication counts, Hirsch's h (using Kelly & Jennion's (2006, p. 167) method), Egghe's g, median citation counts, mean numbers of citations per publication, Bihui's a and the proposed v and h^+ metrics for each researcher were calculated. The results for the metrics were correlated with each other and the population segmented and compared, including by gender, institution, and position held. The skewness metrics, Λ, h_c-range, g_c-range, M_r and h_c, were also correlated with each other. Since some metrics produced discontinuous data, Spearman's Rank Correlation Coefficient was used. The significance, if any, of resulting correlation coefficients were then examined to determine any patterns and establish whether any of these four metrics are strongly correlated with simpler metrics, which might therefore be used instead in future. For the comparison of

metrics, tests of correlation significance were two-tailed. For comparisons of peer ranking with *h*-index scores, tests for significant correlation were one-tailed.

Web of Science failed to record authors' addresses for proceedings, and *Google Scholar* omits authors' addresses for books. For the *Google Scholar* data, a web search was conducted where university affiliations were unclear but even this failed to confirm some researchers' identities. Where the authenticity of a work could not be confirmed, it was included if the subject fitted with the research interests listed on the researcher's university webpage or authorised publication lists. Where these were unobtainable, works were arbitrarily included or excluded on the basis of whether their topics were consistent with the remainder of the researcher's indexed publication output.

No attempt was made to locate the uncited publications listed by *ScHolar Index* for individuals with very long lists of uncited publications, which may have affected the accuracy of the total publication counts, *v*, or overall median citation counts calculated using *ScHolar Index* data for individuals with many uncited publications (almost all highly productive authors). These were again included or excluded on the basis of consistency of topic with other publications by the same author.

Authorised publication lists were not used because those on websites were often out of date, and because it was thought unlikely that all researchers would have provided authorised publication lists.

3.3   Limitations of data collection procedure

3.3.1 Comparison of metrics

ScHolar Index may have exaggerated publication counts by indexing publications and pre-prints separately.

Sheila French was excluded from the analysis because she could only be located through her institution through *Web of Science*. Peter Willett was excluded because his productivity and citedness were both so great that his inclusion would have distorted the results. Replacements for these two individuals were randomly selected from the sample frame to restore the sample size to one hundred.

Patents were excluded because they appeared to be mostly US patents and when examined, many were obvious homographs and all were of uncertain affiliation.

3.3.2 Questionnaire survey

It was not known whether the listings of researchers on all departmental websites were complete; in the course of following up missing email addresses, it became obvious that not all the lists were current. Seven email addresses provided on websites were no longer active, two researchers declined to reply because they were no longer active lecturers, and two because they considered their research to lie outside of Information and Library Science.

Only researchers deemed to be working in the area of Library and Information Science were included in the sampling frame. It was difficult to define the boundary of this field because of inter-disciplinary research and collaboration, and the increasing convergence of different research disciplines. Detailed examination of citation indexes later revealed that some computer scientists and social scientists who carried out research in Library and Information Science had been included.

Several participants commented that the survey questions were difficult (two professors argued that it was impossible) to answer because they required researchers to make generalisations about their publications as a whole. It was expected that individual responses would be biased by the research conducted by participants but because the answers of many researchers were being compared, it was assumed that such biases would cancel one another out, even if a recency effect

was evident and participants' more recent research influenced the answers more than earlier research. All researchers were asked at the same time, however, so it was possible that an external factor, such as a funding initiative, might have systematically influenced the research being conducted. If there was a recency effect, where responses were more influenced by participants' recent publications, then systematic effects on publication type might have systematically biased the results.

Some survey questions were criticised for having a slight positivist bias.

Respondents could not have been asked to reply by return of email to be removed from the mailing list because this could have been considered SPAM. Even the first mailing of the survey was classified as SPAM by Northumbria University and the originating email address blocked. Several out-of-office replies were also received, suggesting that some participants were away until early July.

Some teaching staff had never published research.

Although the networked version of Internet Explorer 6 on the Loughborough University network allowed previous answers to be changed, it was reported that it was not possible to alter previous questions using the same application on other university networks. This was a potential benefit because it may have prevented answers being altered in the light of subsequent questions.

3.3.3 Comparisons of databases and genders

Some metrics yield discrete data, where the step size was large compared to the standard deviation. Such distributions do not resemble the continuous Normal distribution and are therefore unsuitable for parametric analysis (Bowker & Randerson 2005, p. 16). The databases and metrics were therefore compared using a Sheirer-Ray-Hare test.

Most metrics were discontinuous and therefore violated an assumption of ANOVA (Bowker & Randerson 2005, p. 16). All two-factor non-parametric tests require a balanced design. There were different numbers of institutions of different RAE (2001) grade, so a comparison by RAE score was not possible. Only one male researcher was sampled from London Metropolitan University, so Friedman's test (Siegel & Castellan 1988, p. 174) was used to compare the academic ranks of female and male researchers in different departments. One male and one female researcher was randomly selected from each institution and their ranked academic positions compared between different institutions.

3.4 Comparison of peer review and citation metrics

Researchers were asked to name and rank the five best researchers in their department in order of decreasing research performance. If more individuals had been ranked, the critical value of r_S drops sharply but since one department only comprised seven researchers, only five names were requested. Hirsch's h-index was calculated for different researchers and used to rank the individuals selected. Each set of h-index derived rankings and peer-rankings were then compared using Spearman's rank correlation.

The proportion of researchers of each gender ranked in the top five for each department by their peers was compared with the proportion of each gender in the sample frame.

3.5 Qualitative analysis

The respondents' natural language was used to code the data initially (Heath & Cowley 2004, p. 144), and the application of controlled researcher codes left until as late as possible.

Memoranda were recorded on notelets, as suggested by Heath & Cowley (2004, p. 144), and a reflexive journal (Gorman & Clayton 2005, p. 210) was kept throughout to assist consideration of emerging concepts without reference to the raw data (Gorman & Clayton 2005, p. 218), to develop ideas and analytical strategies with respect to the rest of the data gathered.

3.6   Factor analysis

3.6.1 Number of significant factors

Factors with eigenvalues greater than unity were included in the analysis.

3.6.2 Factor rotation

The significant factors detected were successfully rotated to simple structure using the orthogonal *Varimax* rotation with Kaiser normalisation. Other rotations were checked to ensure no more parsimonious solution was possible.

3.6.3 Identification of significant factors

Due to a poor response rate, factor loadings that exceeded 0.5 in magnitude were regarded as significant, as proposed by Manly (1994, p. 101). The correlation matrix and mean question scores were critically studied and related to the factors.

4   Results

4.1   Population and sample demographics

Figure 2 represents the proportion of male and female researchers in the underlying population and in the portion of the random sample who were also indexed by the two citation databases. Only 77 of 100 randomly sampled researchers were indexed by *Web of Science* and 86 by *Google Scholar* (when accessed through the *ScHolar Index* application).

Population of UK Library and Information Science researchers (N=266)

Gender
- Female
- Male

Web of Science (N=76)   *ScHolar Index* (N=86)
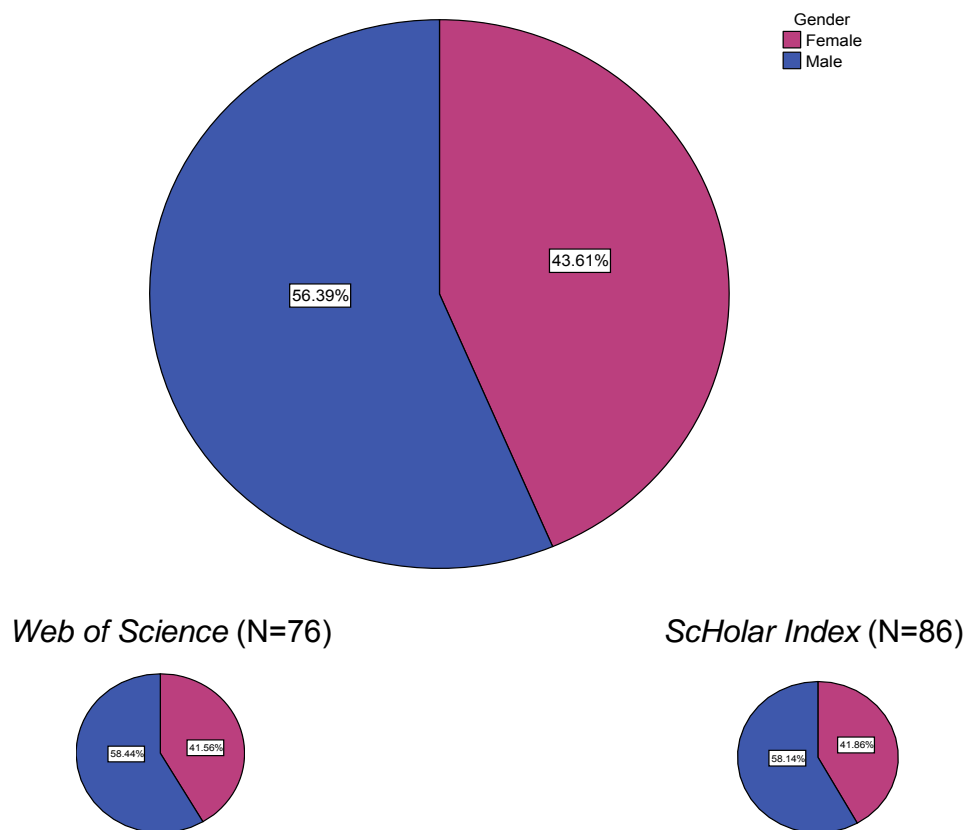
Figure 2. Comparative pie charts representing breakdown of population and those indexed by the different databases by gender

Figure 3 shows the proportion of academics of different academic rank in the parent population, and in the data gathered using the two citation indexes. After *Web of Science* (*WoS*) had been updated, it was realised that the total number of publications had been incorrectly recorded for Lucy Tedd, who was consequently dropped from the analysis so that the same individuals were compared for all metrics, and were therefore directly comparable.

The University of Northumbria, Brighton University and the University of the West of England failed to specify the ranks of most academics. Their staff were excluded from all comparisons of different academic ranks, together with researchers described only as "course co-ordinators" because the academic status of such individuals might have varied between institutions.

The descriptions of academic positions varied so for ease of analysis individuals were ascribed to arbitrary broad bands of approximate prestige based upon the descriptions of their grade given on their university websites, which are henceforth referred to by the descriptors given in Table 1. Each band was numbered and these numbers were ranked and used in the subsequent analyses.

Table 1. Standardised descriptions and numeric codes for academic ranks

Position(s)	Description of academic rank	Numerical rank
Post-doc/Tutor/Research Associate/ Research Assistant/Research Officer	Post-doc/Tutor	1
Lecturer/Researcher/Senior Lecturer (at universities with Principal Lecturers)	Lecturer	2
Senior lecturer (at universities lacking Principal Lecturers)/Principal Lecturer/Research Fellow	Senior Lecturer	3
Reader/Senior Research Fellow	Reader	4
Professor	Professor	5
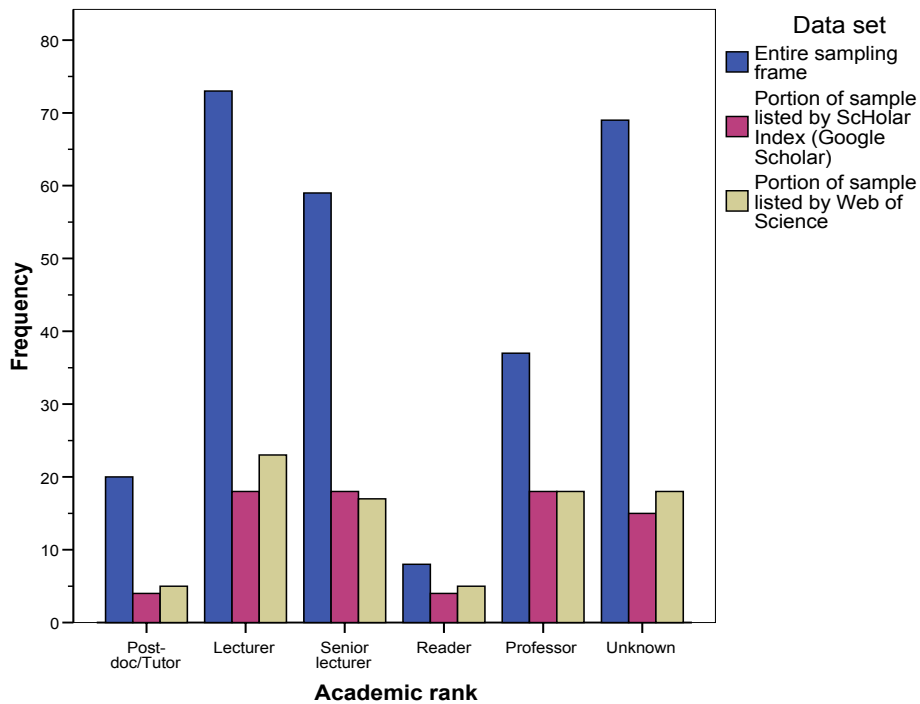
Figure 3. Comparative bar chart representing breakdown of population and sample by broad academic rank

Table 2 shows the breakdown of the population and data gathered from the two citation indexes by academic institution.

Figures 1-2 and Table 2 show that, although limited by the coverage of the two citation indexes, the data gathered was still representative of the underlying population with respect to gender balance, academic rank and institutional affiliation.

Table 2. Demographic breakdown of population and samples by institution

Database	Gender	Institution															
		RGU	City	Leeds Metropolitan	Liverpool	London Metropolitan	Loughborough	Manchester Metropolitan	Napier	UCL	Strathclyde	Brighton	Northumbria	Sheffield	Aberystwyth	UWE	Total
WoS	Female	5	0	1	1	2	4	2	1	2	1	1	4	5	2	0	31
	Male	2	2	1	3	1	5	2	1	3	4	4	3	7	3	4	45
	Sub-total	7	2	2	4	3	9	4	2	5	5	5	7	12	5	4	76
ScHolar Index	Female	4	0	1	0	2	4	4	1	2	2	2	5	5	3	0	35
	Male	3	2	3	3	1	5	3	1	3	4	3	5	7	4	4	51
	Sub-total	7	2	4	3	3	9	7	2	5	6	5	10	12	7	4	86
Overall	Female	10	5	2	4	6	12	15	2	7	6	5	20	12	10	0	116
	Male	10	10	7	7	1	16	8	7	11	11	9	14	18	10	11	150
	Total	20	15	9	11	7	28	23	9	18	17	14	34	30	20	11	**266**

4.2 Correlation analysis of calculated metrics

All the indices calculated were correlated with one another for each citation database independently (Appendix G). Annual citation rates were only available from *Web of Science*.

Most correlations were positive and very highly significant (P<0.001) for both the *Web of Science* and *ScHolar Index* data, with the exceptions of *h* and Λ, which were not significantly correlated (r_S=0.229; P=0.073 (3 d.p.). Unless stated, all test statistics and P-values are given correct to three decimal places.

4.2.1 *h*-index correlations

WoS data

ScHolar Index data
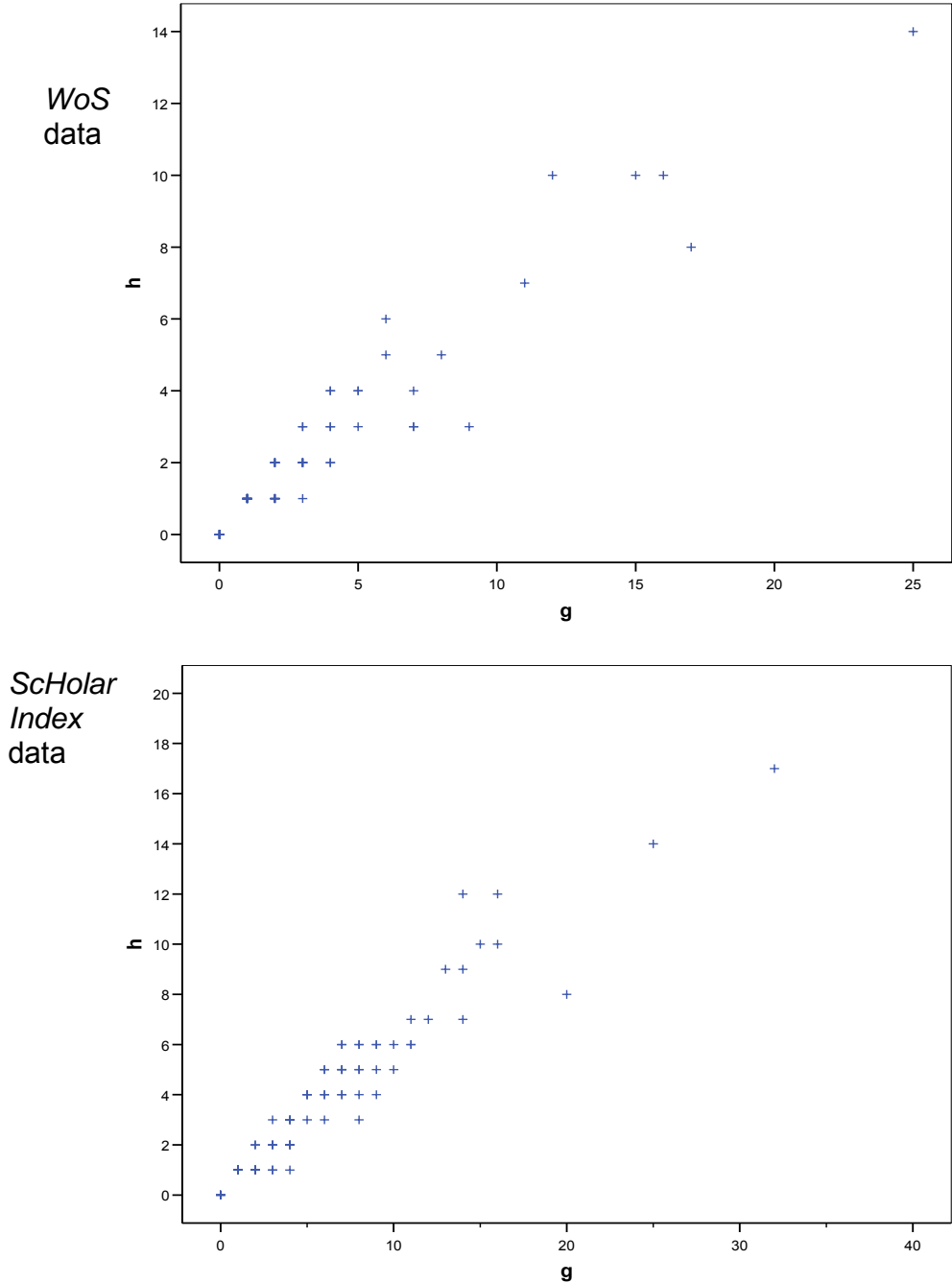
Figure 4. Scatter plots of *h* against *g*

Figure 4 shows that *h* and *g* were strongly, linearly and positively correlated.
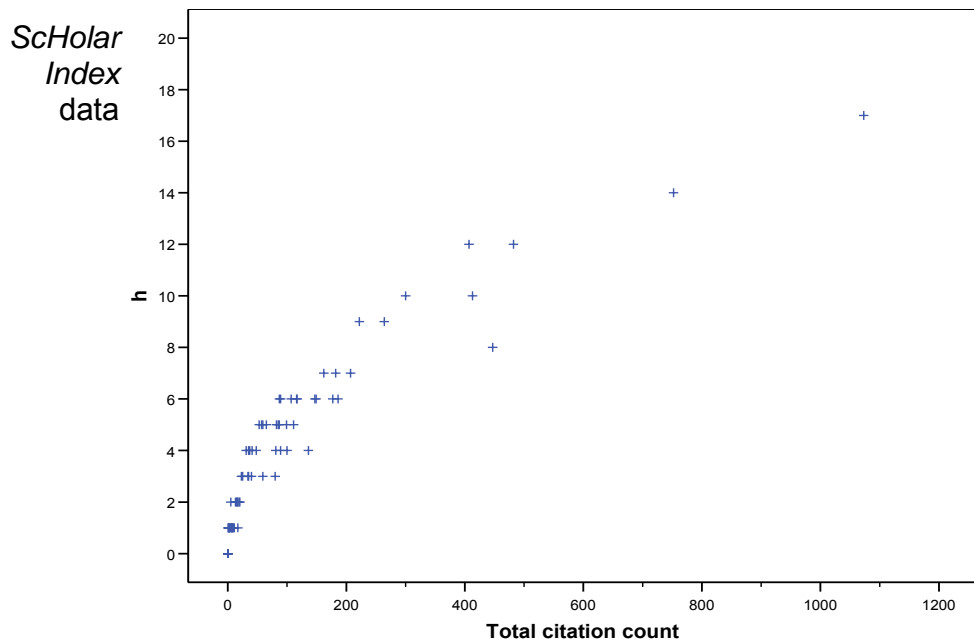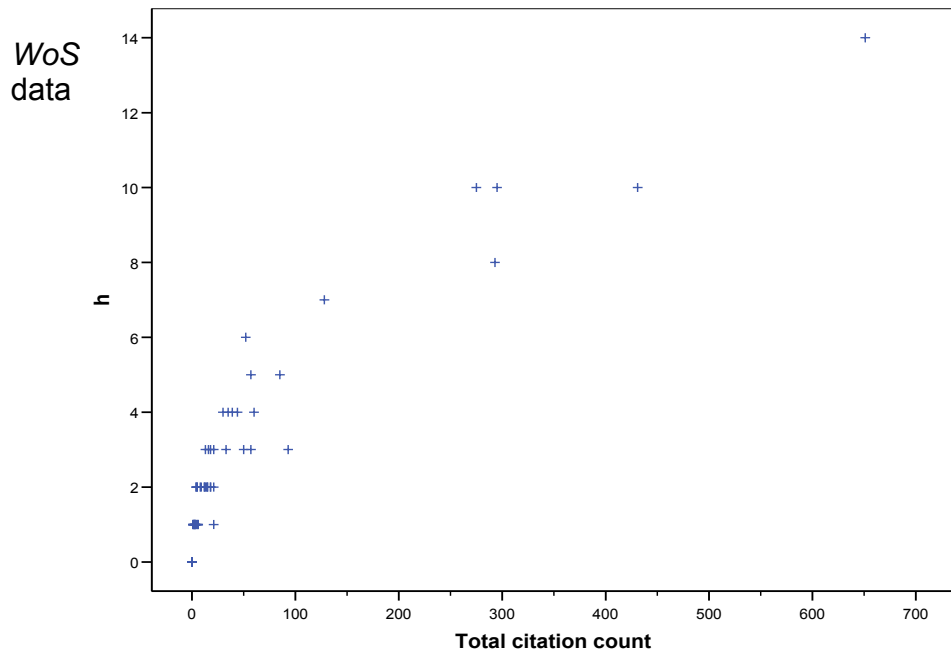
Figure 5. Scatter plots of *h* against total citation count

Figure 5 shows that *h* correlated strongly and curvilinearly with total citation count. The rate of increase in *h* decreased with increasing total citation count, such that $h \approx \sqrt{\text{(total citation count)}}$.
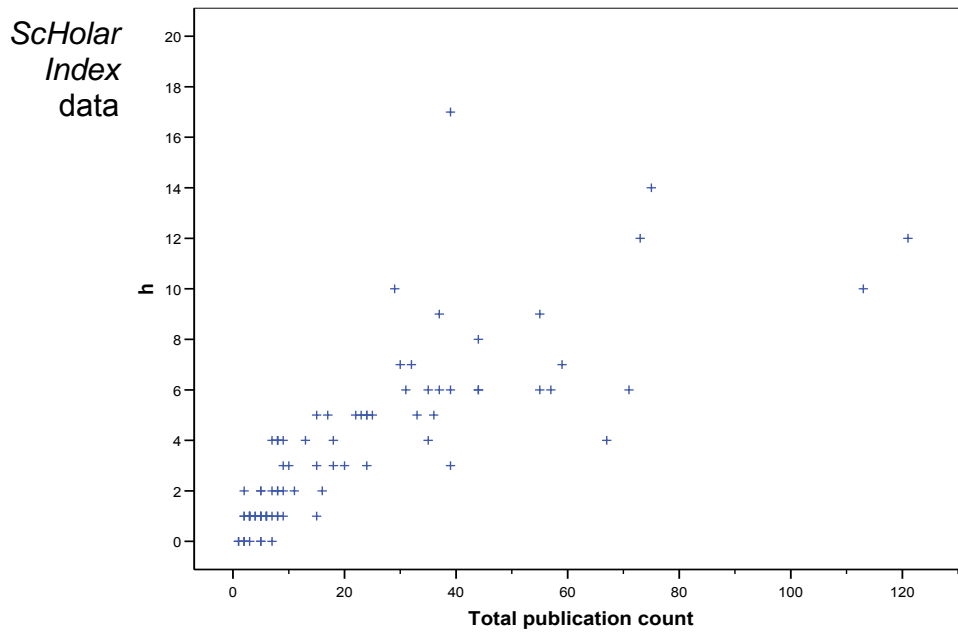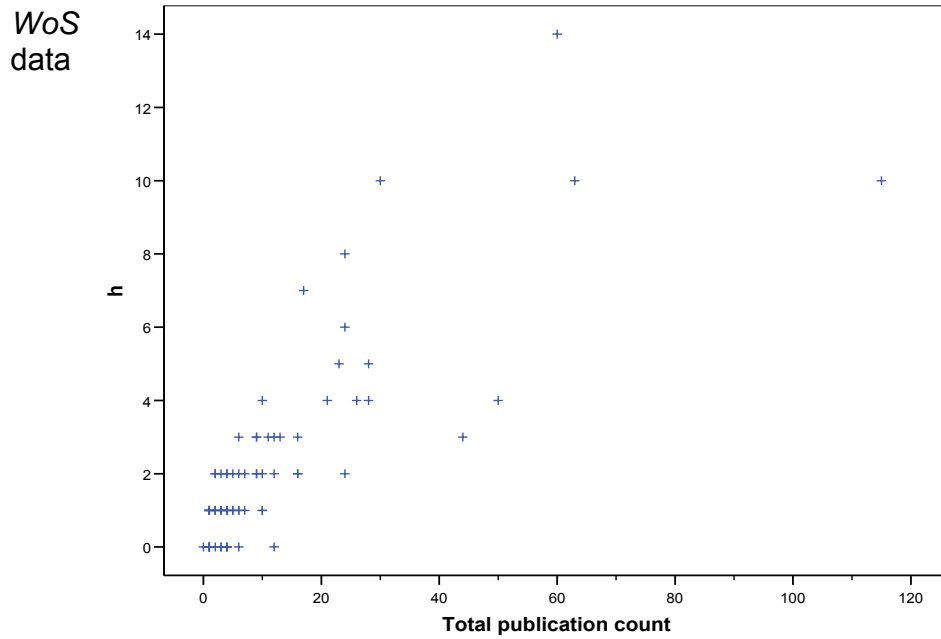
Figure 6. Scatter plots of *h* against total publication count

Although *h* and total publication counts were strongly correlated, Figure 6 shows scatter increases with increasing total publication count, with outliers appearing at low total publication counts, below forty publications. The relationship also curves slightly, such that $h \approx \sqrt{(\text{total publication count})}$.
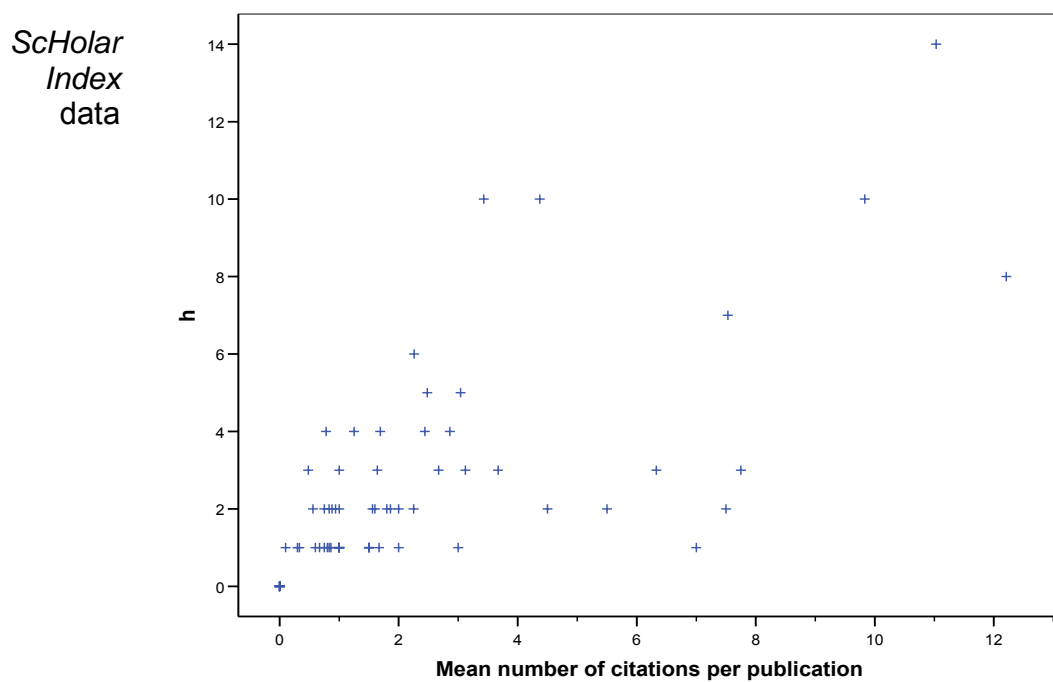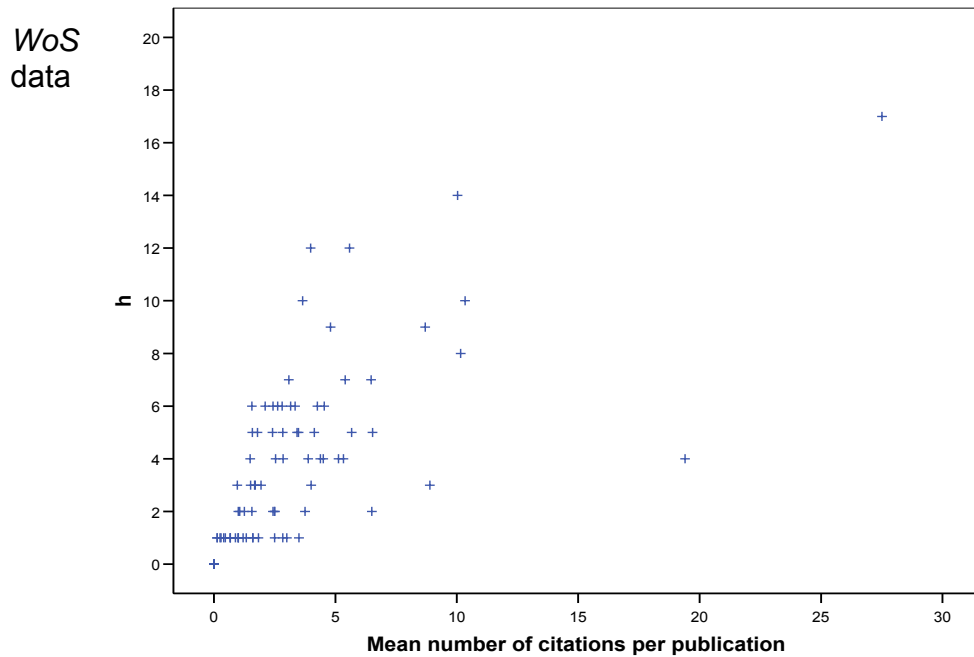
Figure 7. Scatter plots of *h* against mean numbers of citations per publication

Figure 7 shows that scatter was evident at all levels of citedness but increased with increasing citation rate.
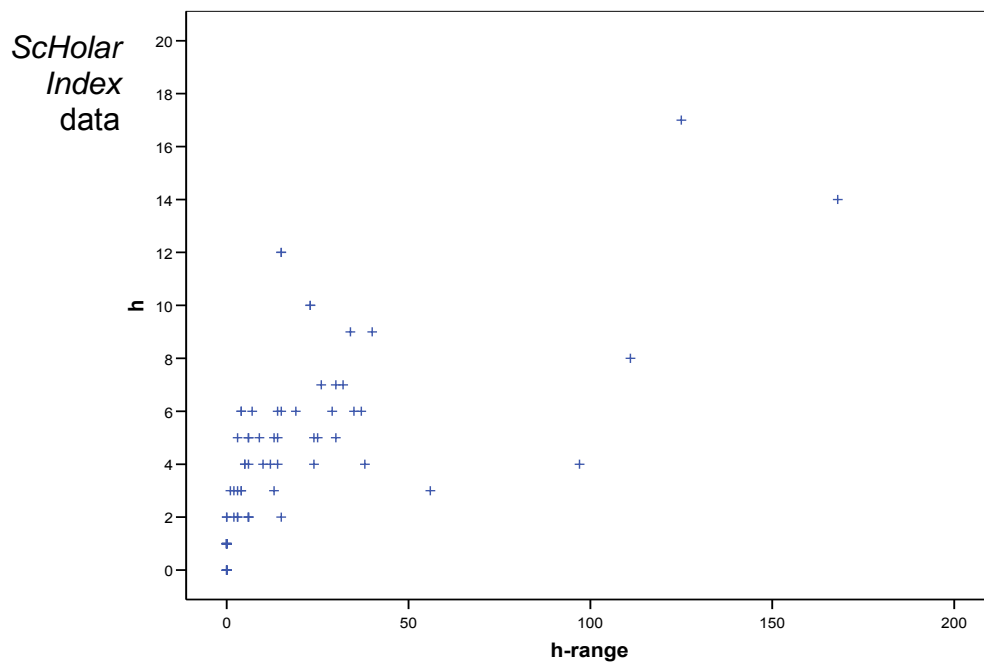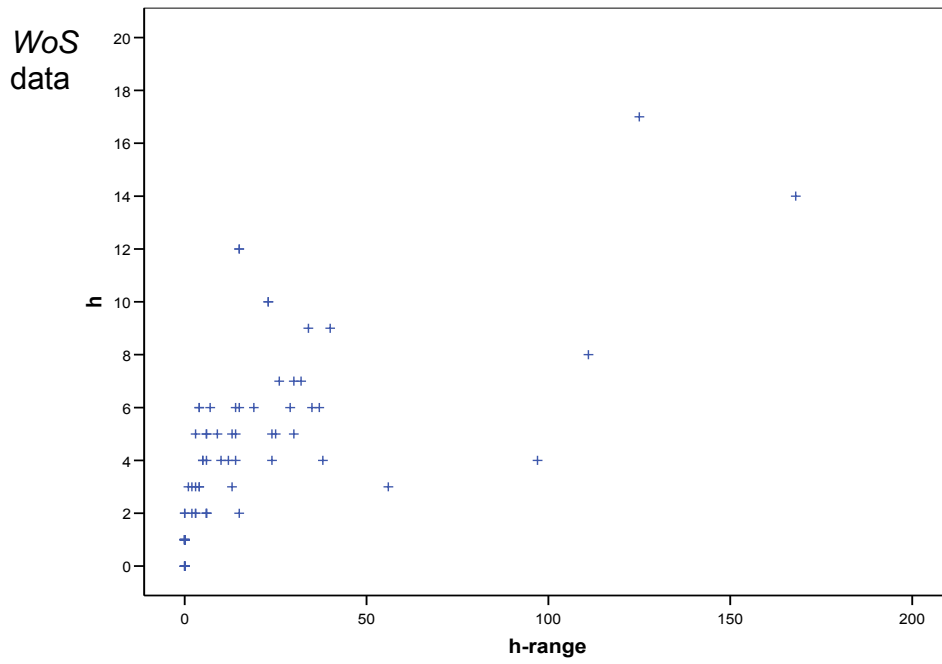
Figure 8. Scatter plots of *h* against *h*-range
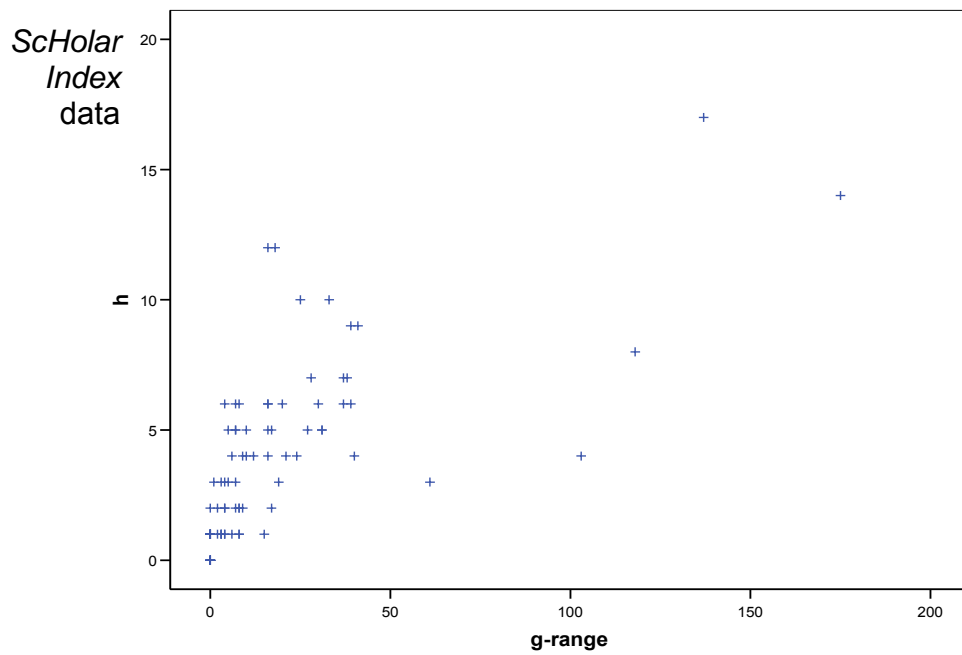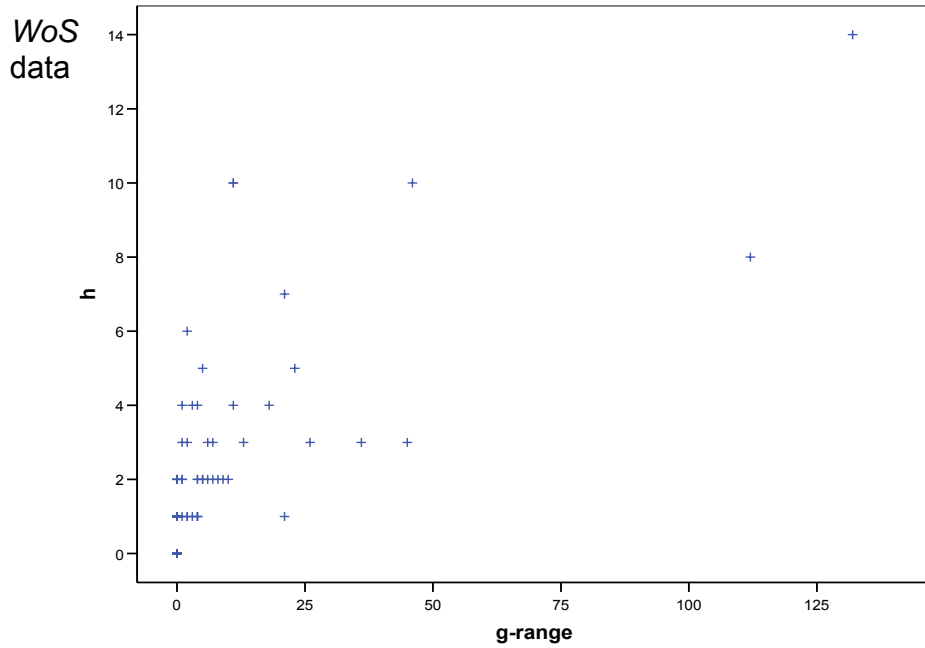
Figure 9. Scatter plots of *h* against *g*-range

Figures 8 and 9 show a general positive association but the degree of scatter was large, especially at low *g*-ranges.
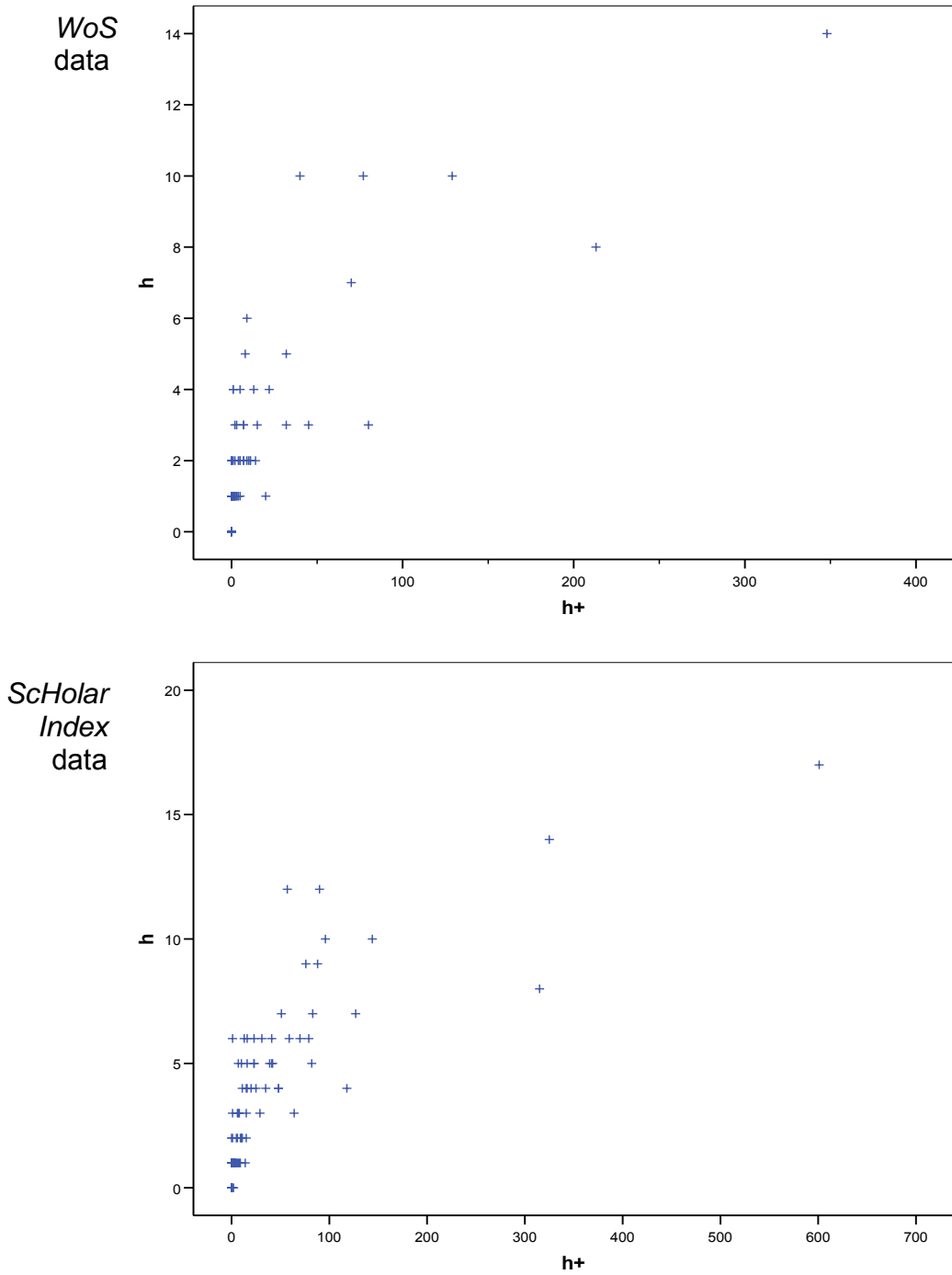
Figure 10. Scatter plots of h against h^+

Figure 10 demonstrates that all the correlations are biased by the clustering of points toward the origin caused by most researchers achieving low h-index scores.

Scattering of data points made it unclear whether h and h^+ scores were related linearly or curvilinearly.
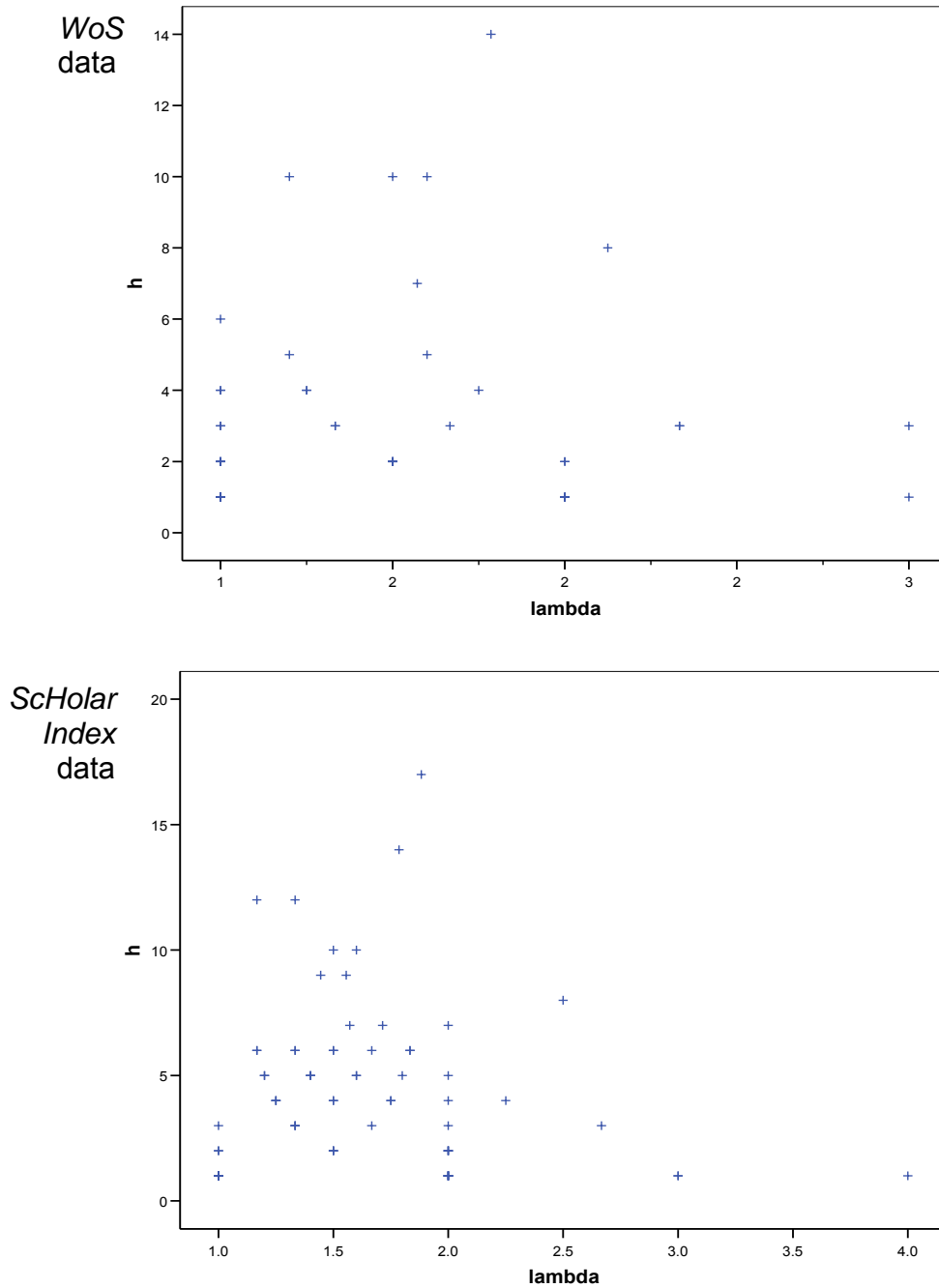
Figure 11. Scatter plots of h against Λ

Figure 11 shows no predictable association between h-index and Λ scores.
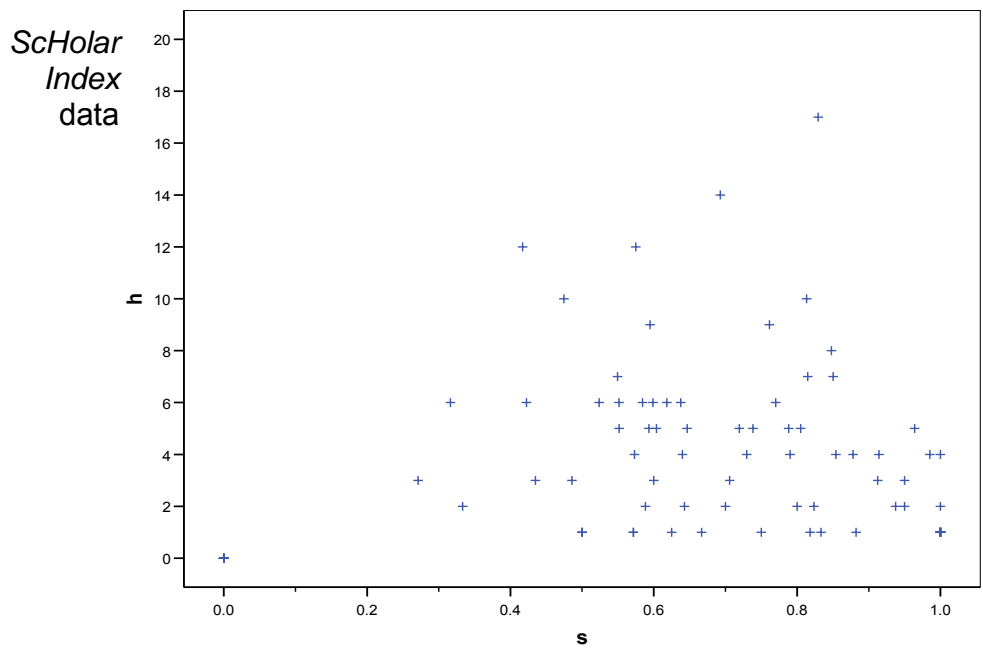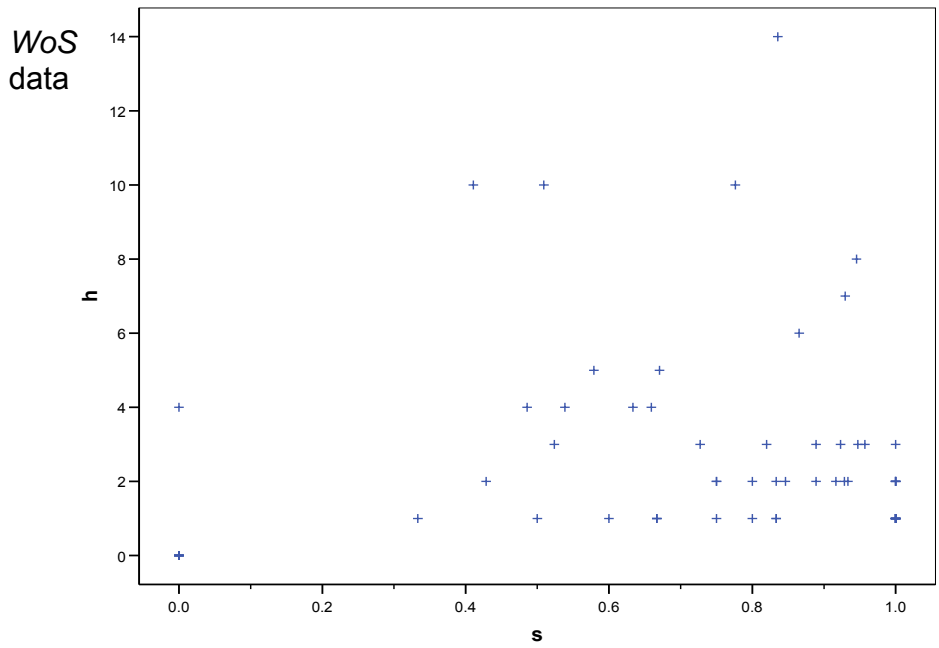
Figure 12. Scatter plots of *h* against *s*

Figure 12 shows no clear association between *h* and *s*, despite their significant and positive correlation (r_S=0.273; P=0.016).
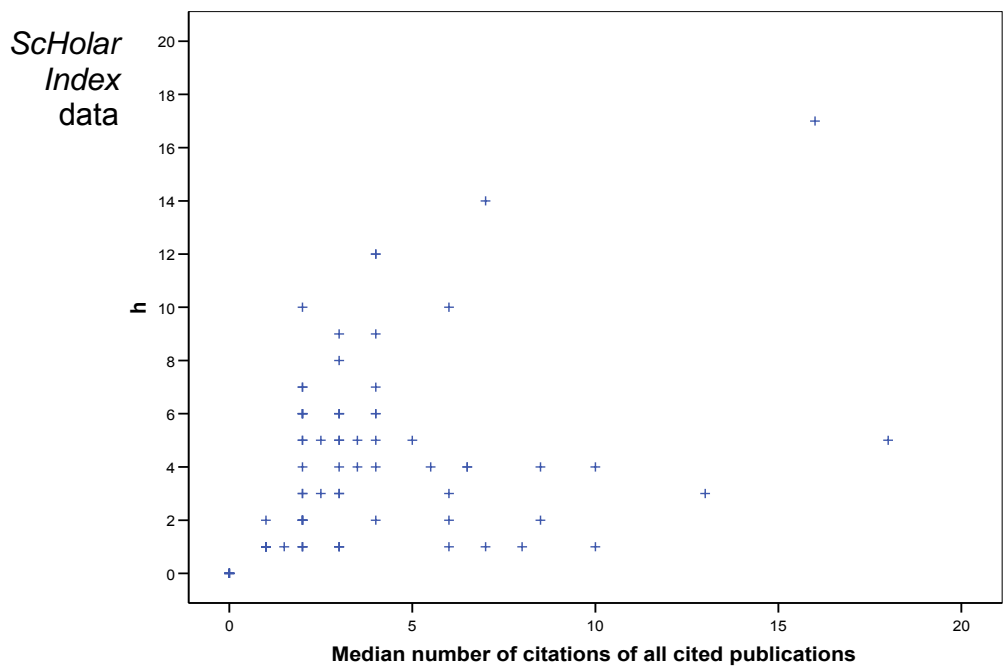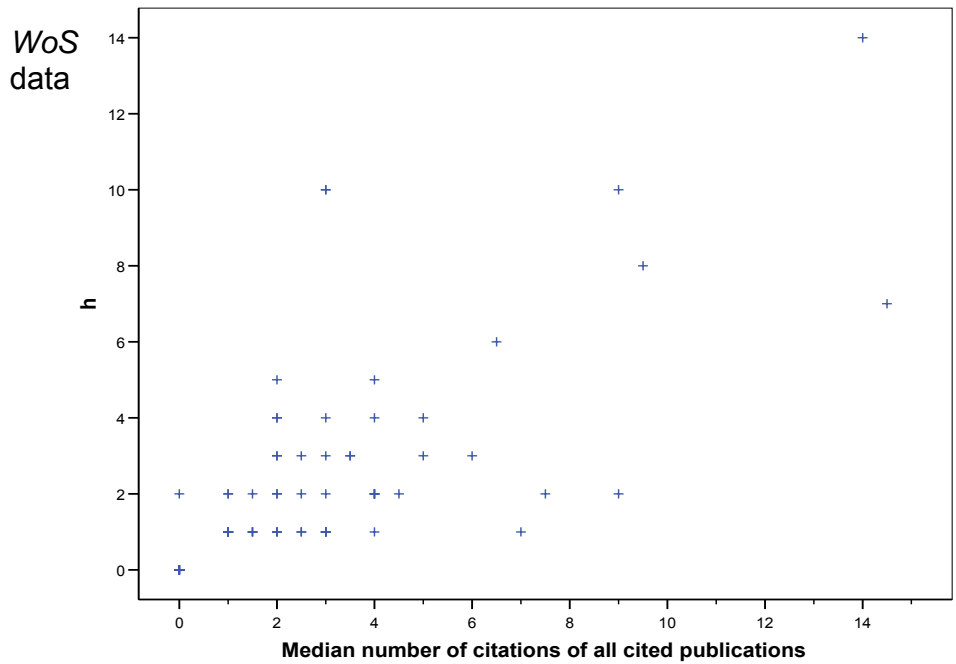
Figure 13. Scatter plots of *h* against median numbers of citations of all cited publications
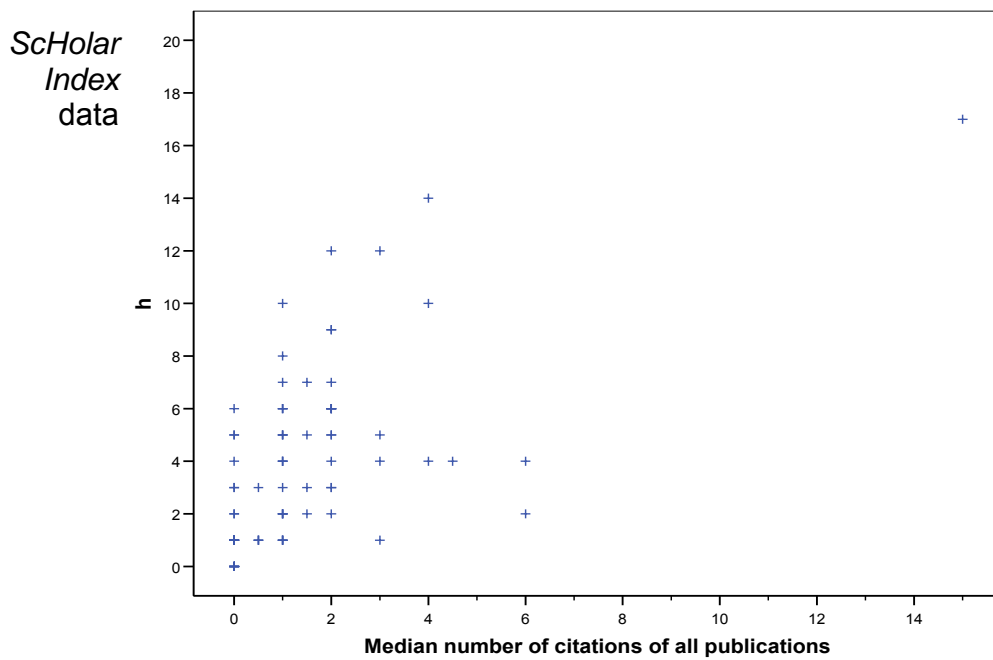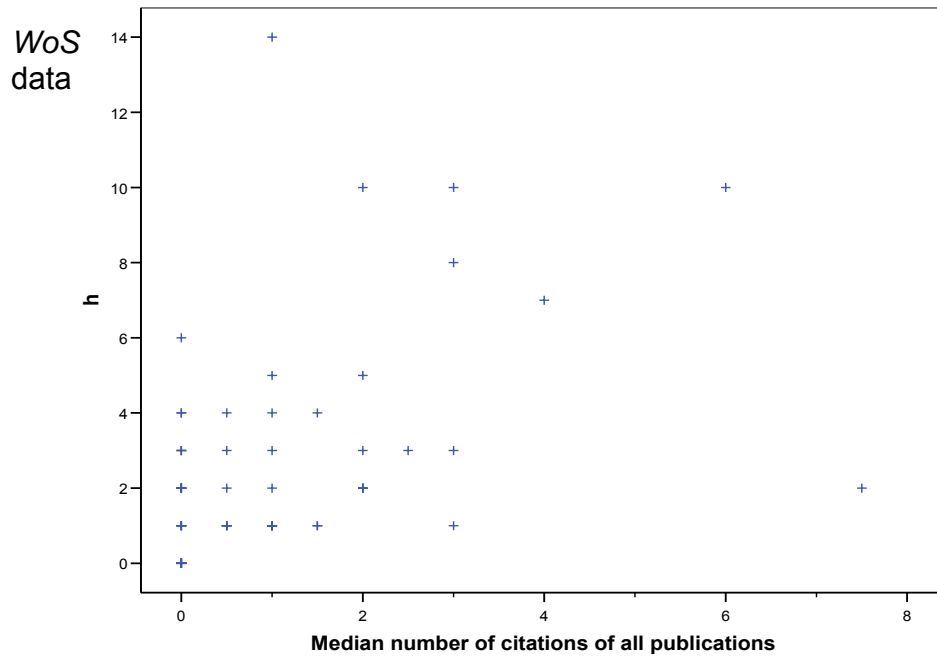
Figure 14. Scatter plots of *h* against median numbers of citations of all publications

Figures 13 and 14 shows that, however calculated, median citation counts were weakly correlated with *h*.
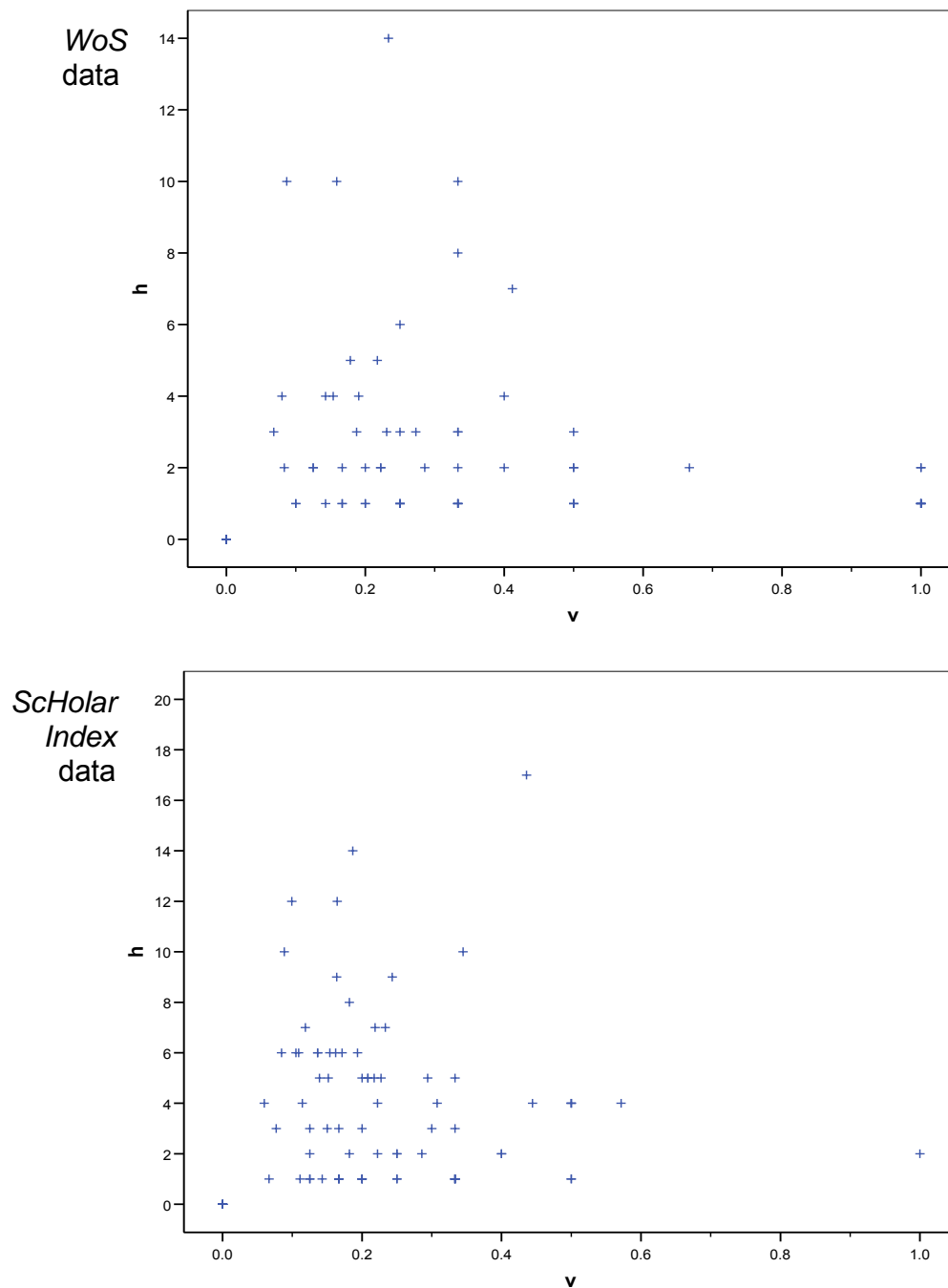
Figure 15. Scatter plots of *h* against *v*

Figure 15 shows no clear pattern in the association between *h* and *v*, and the correlation coefficient was low, but the correlation was still very highly significant (r_S=0.344; P=0.002). Like *h* and *s, h* and *v* were not associated in a predictable manner.
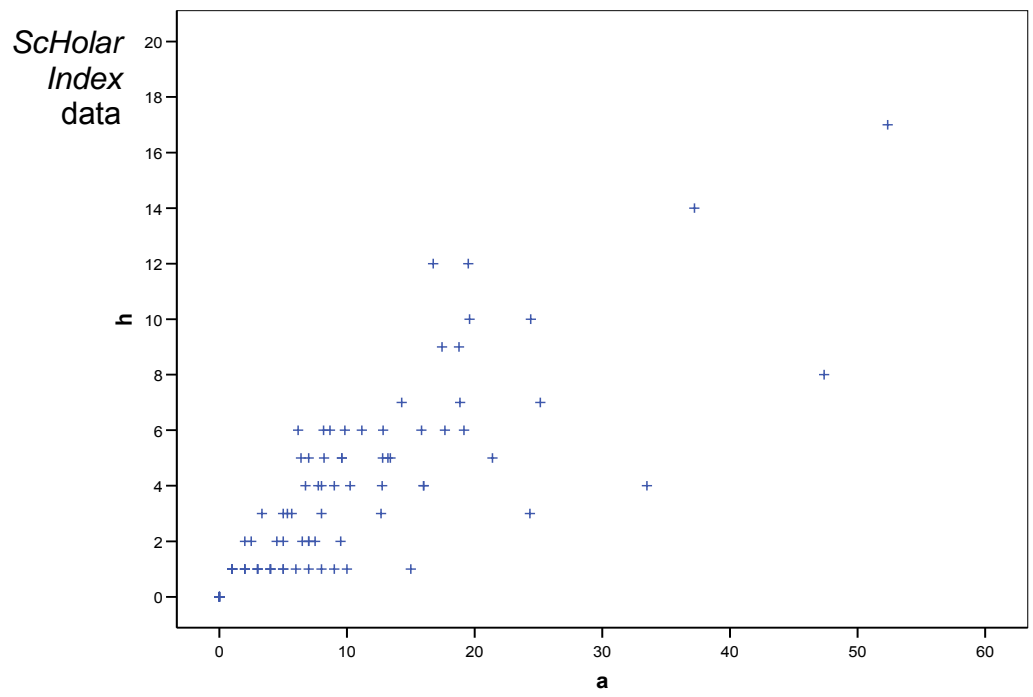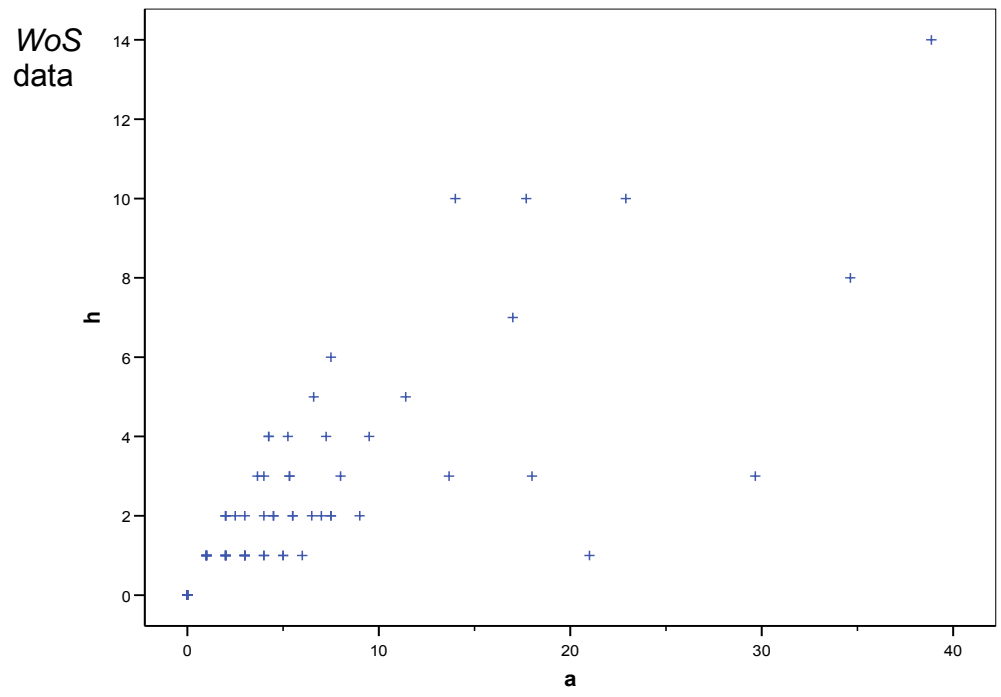
Figure 16. Scatter plots of *h* against *a*

Figure 16 shows a positive, linear association between *h* and *a*.

WoS data

ScHolar Index data
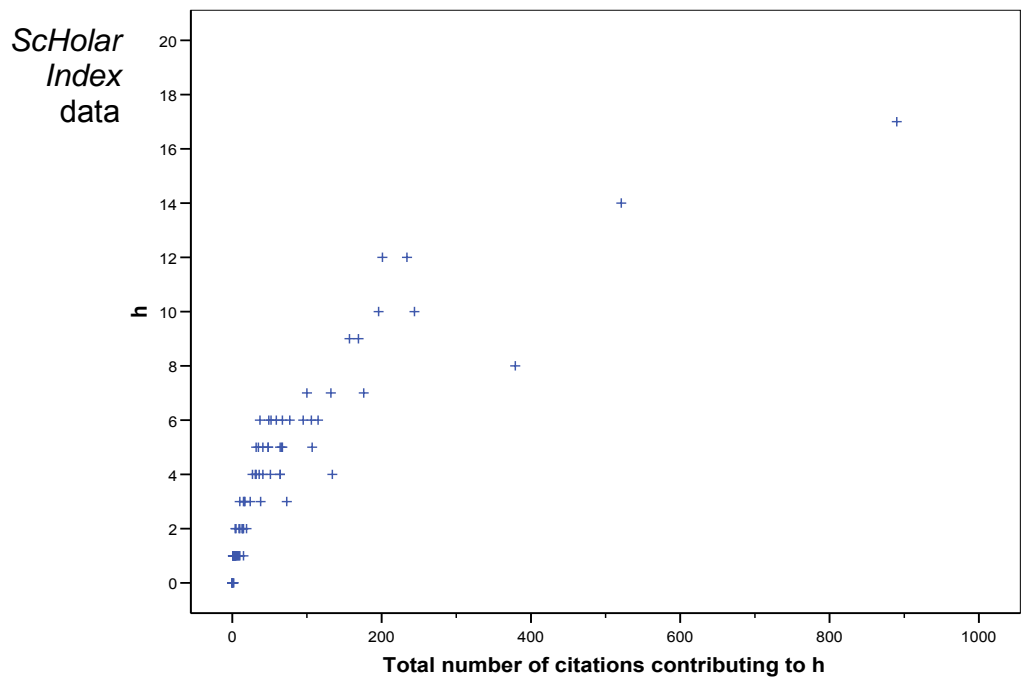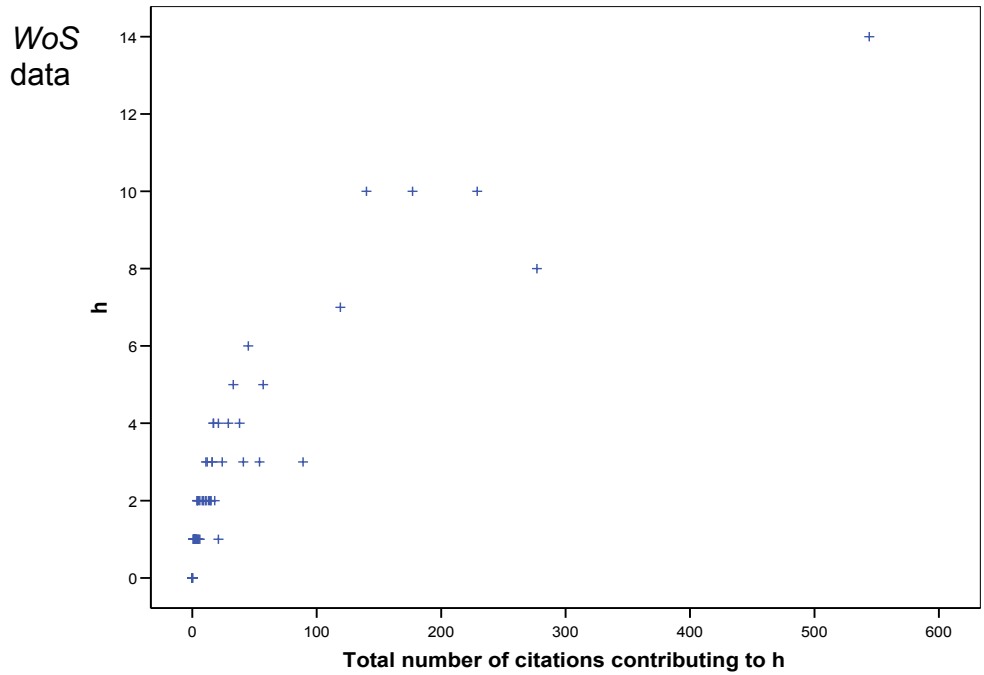
Figure 17. Scatter plots of *h* against the number of citations contributing to *h*

Figure 17 shows a distinct curvilinear (between quadratic and log-normal) relationship between *h* and the total number of citations contributing to *h*.

4.2.2 Correlations of other metrics

WoS data
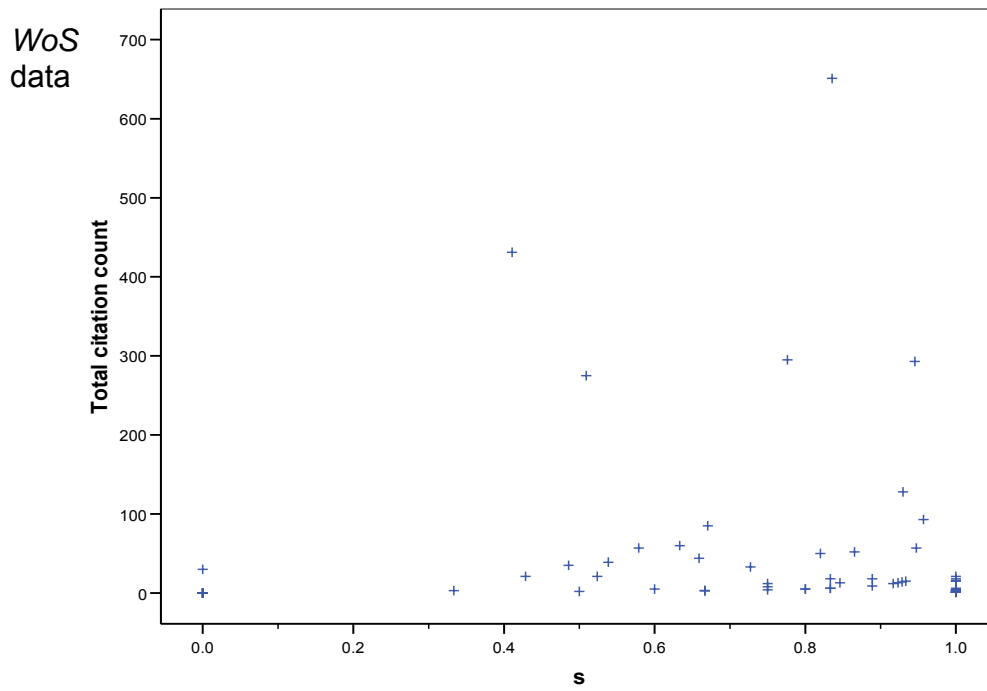
Figure 18. Scatter plot of total citation count against *s*

Figure 18 confirms the lack of any association between total citation count and *s*.
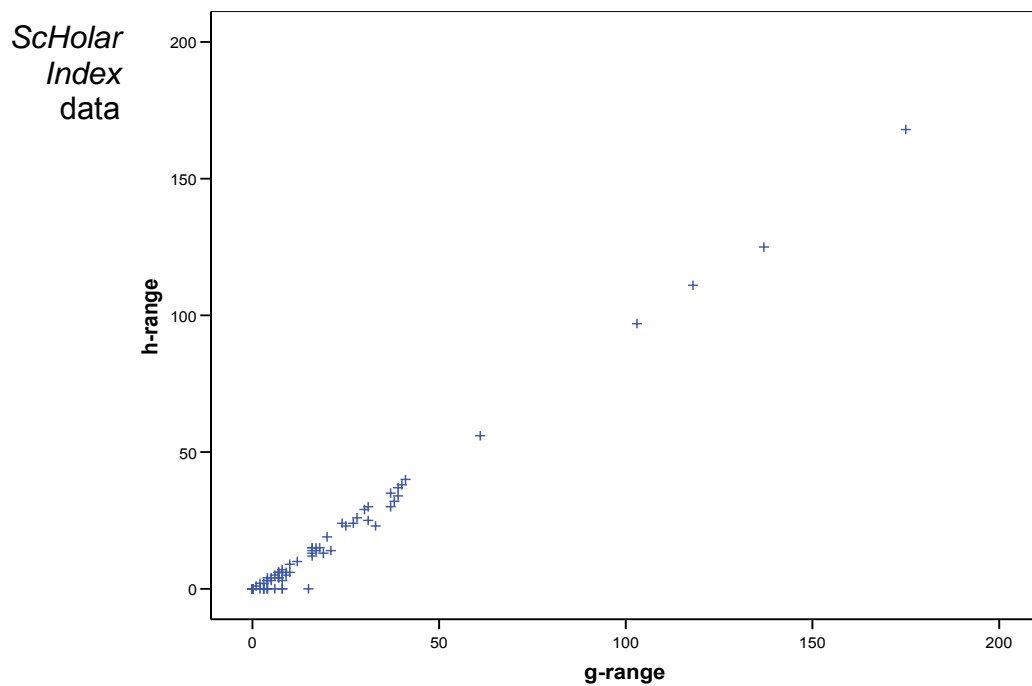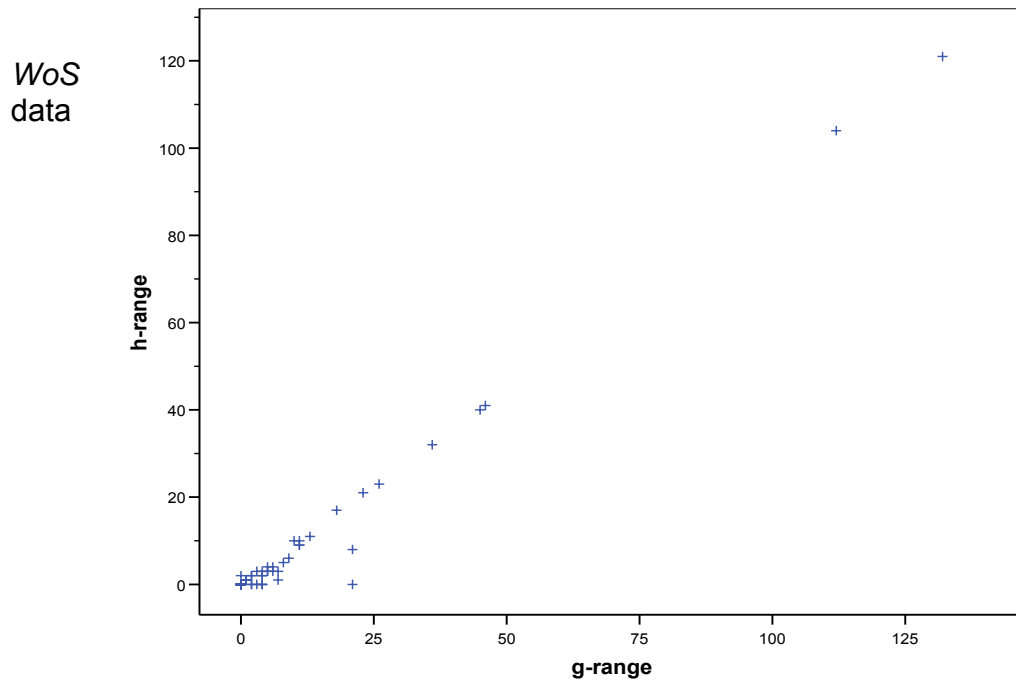
Figure 19. Scatter plots of *h*-range against *g*-range for *WoS* and *ScHolar Index* data, respectively

Figure 19 shows that *h*-ranges and *g*-ranges showed almost perfect positive linear correlation.
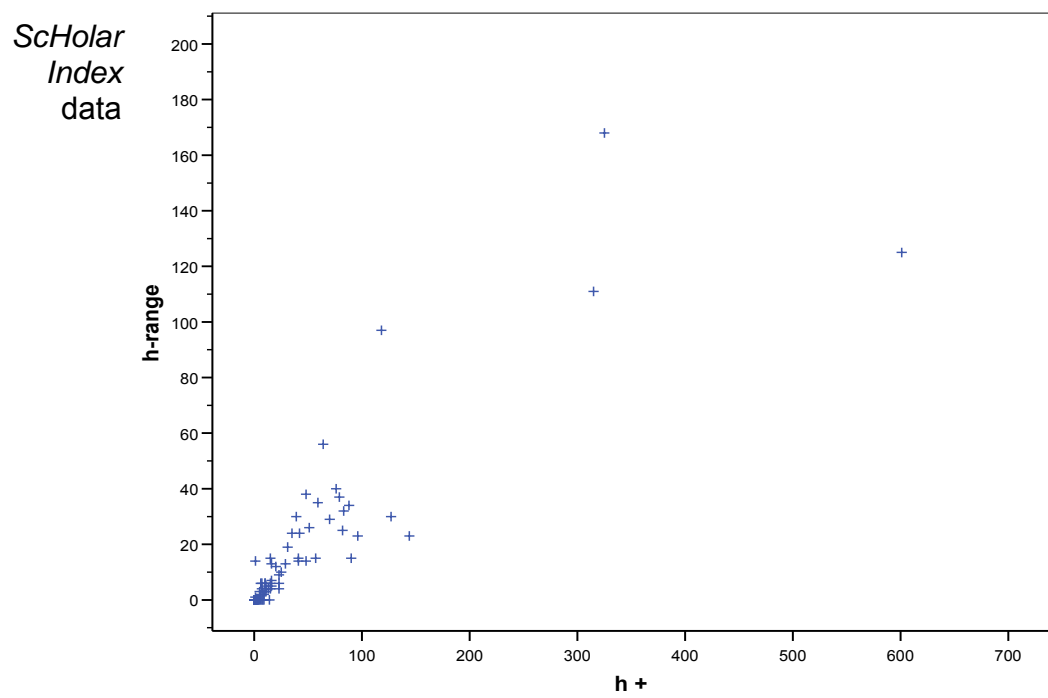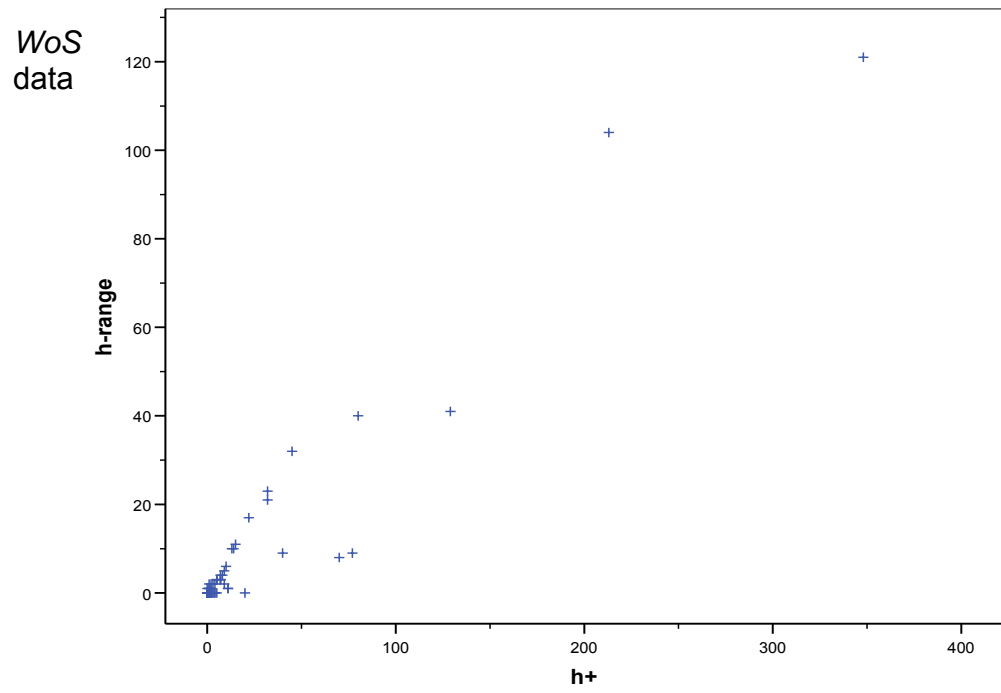
Figure 20. Scatter plots of *h*-range against h^+

Figure 20 suggests that *h*-range and h^+ were correlated linearly with little scatter.
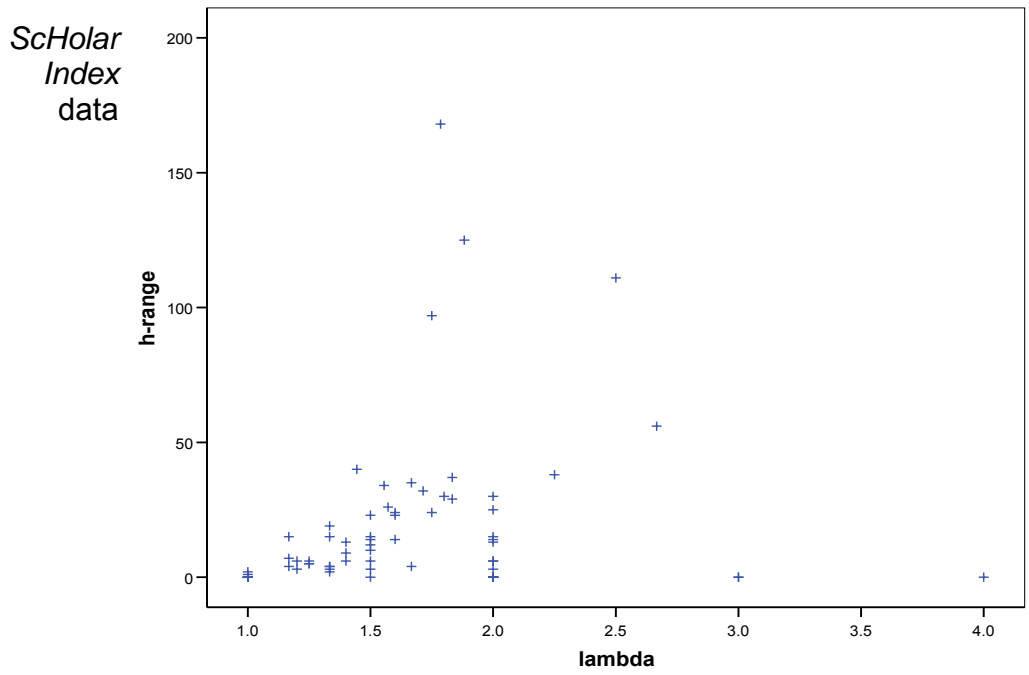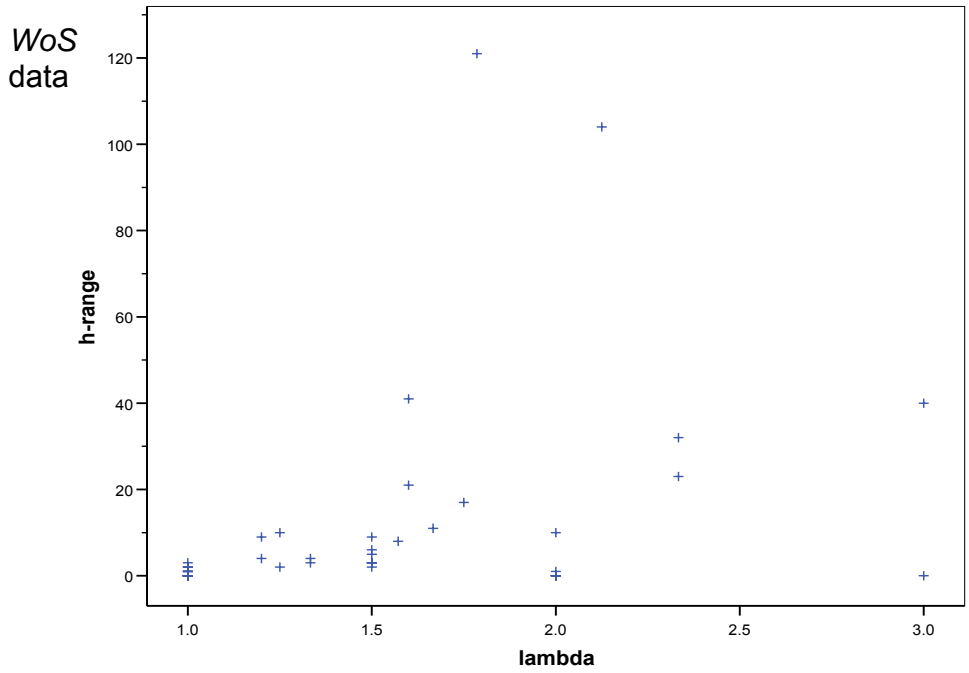
Figure 21. Scatter plots of *h*-range against Λ

Figure 21 shows that *h*-range increased linearly with increasing Λ but that there was much scatter.
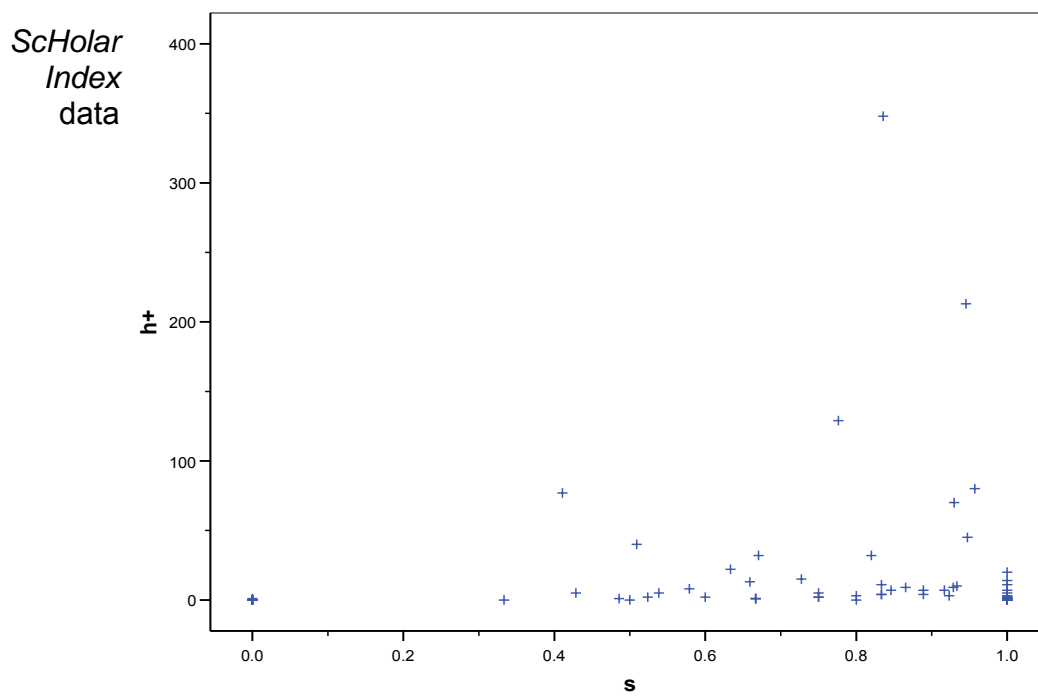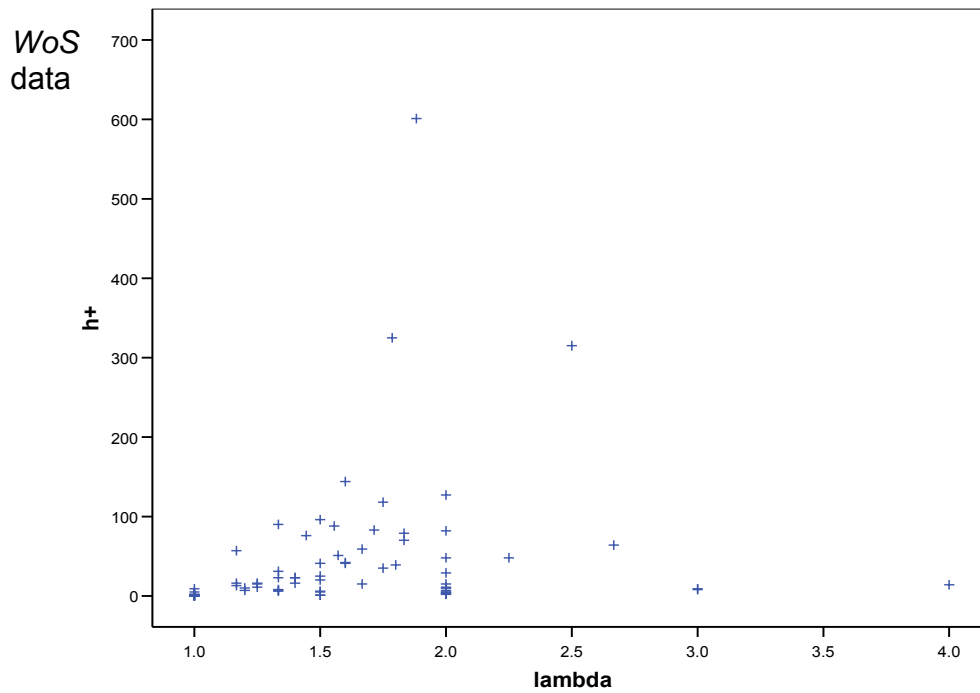
Figure 22. Scatter plots of h^+ against Λ

Figure 22 shows that although strongly correlated ($r_S=0.708$; $P<0.001$), h^+ varied unpredictably with increasing Λ.
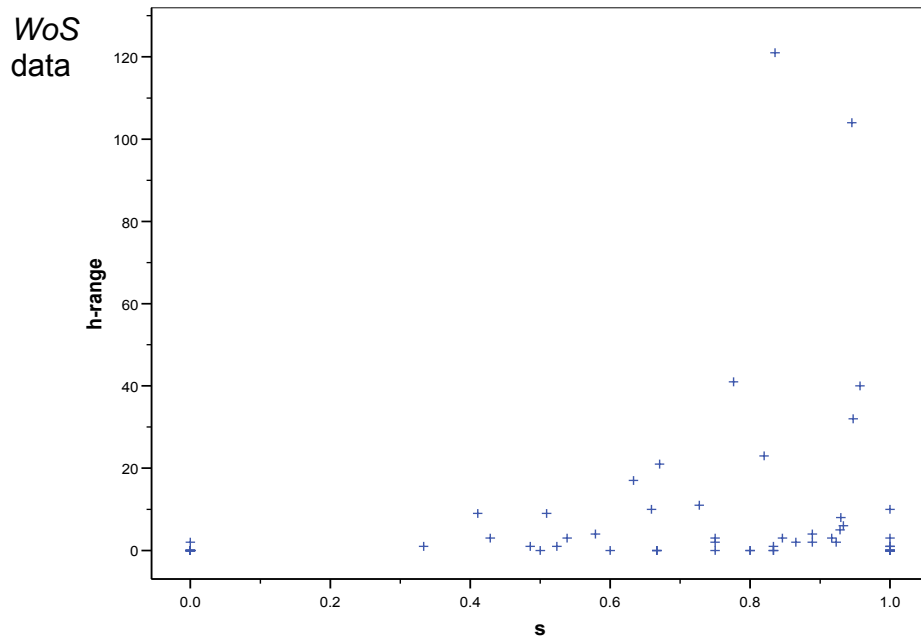
Figure 23. Scatter plot of *h*-range against *s*

Figures 21 and 23 confirm that like *h, h*-range was not correlated with Λ or *s*.
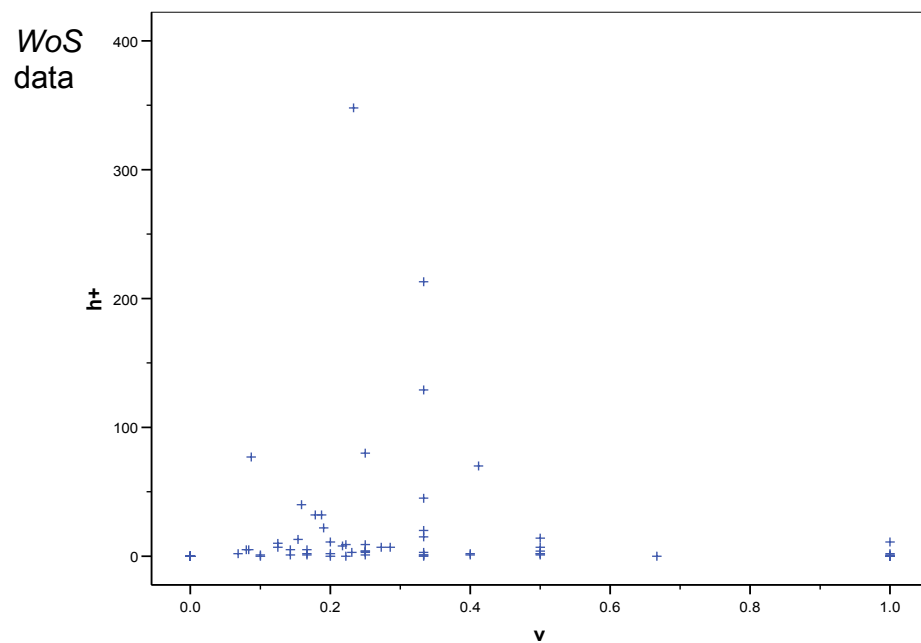
Figure 24. Scatter plot of *h*⁺ against *v*

Although correlated, Figure 24 showed that it was not possible to reliably predict values of h^+ from corresponding values of v.
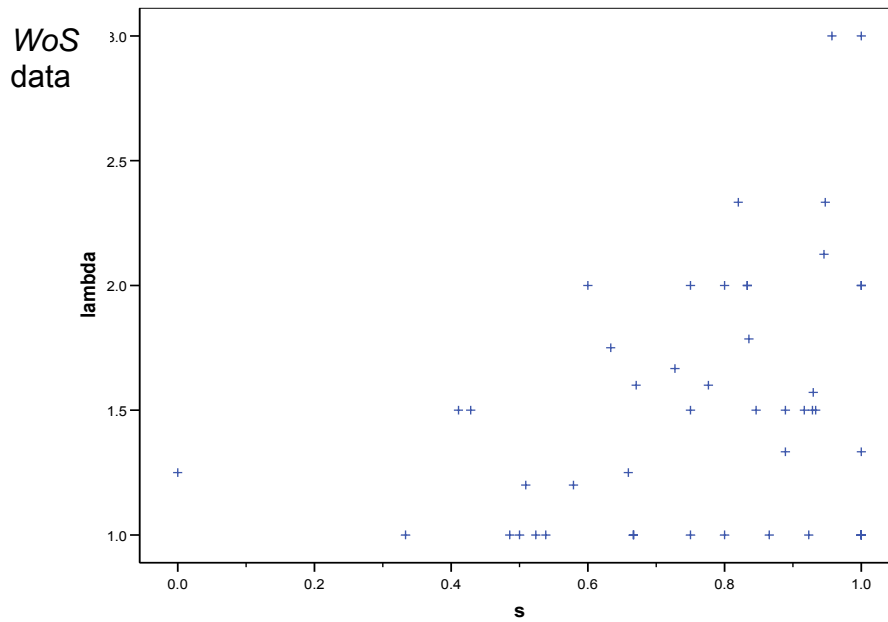
Figure 25. Scatter plot of Λ against s

Figure 25 confirms that Λ was not correlated with s.
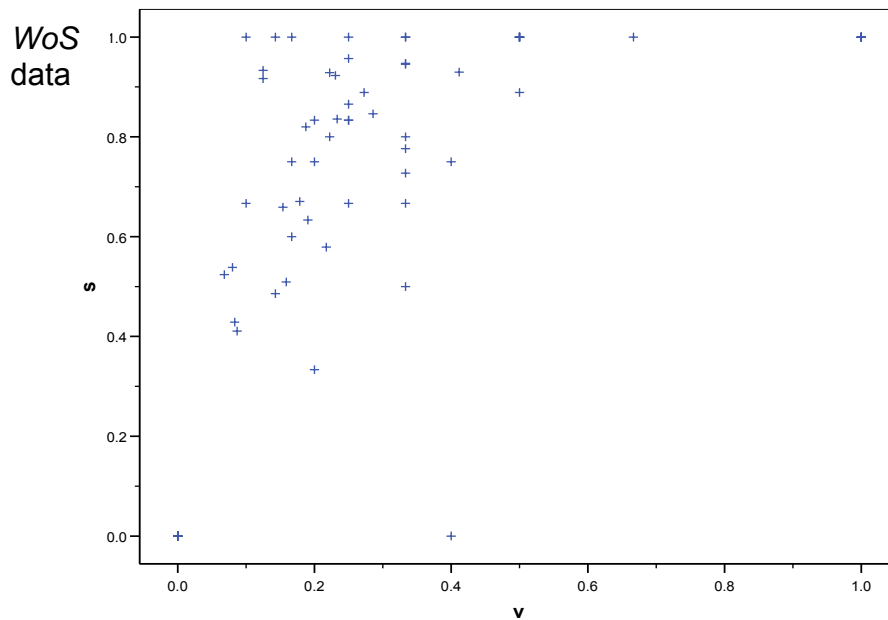
Figure 26. Scatter plot of s against v

Figure 26 confirms that *s* and *v* are positively associated but that the relationship was unclear because of the large amount of scatter.

WoS data

ScHolar Index data
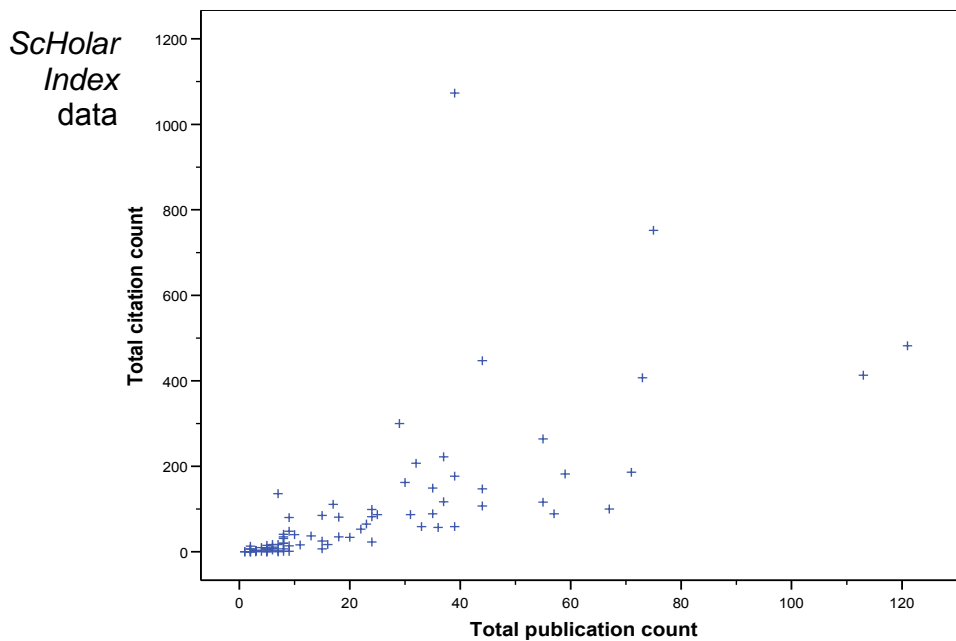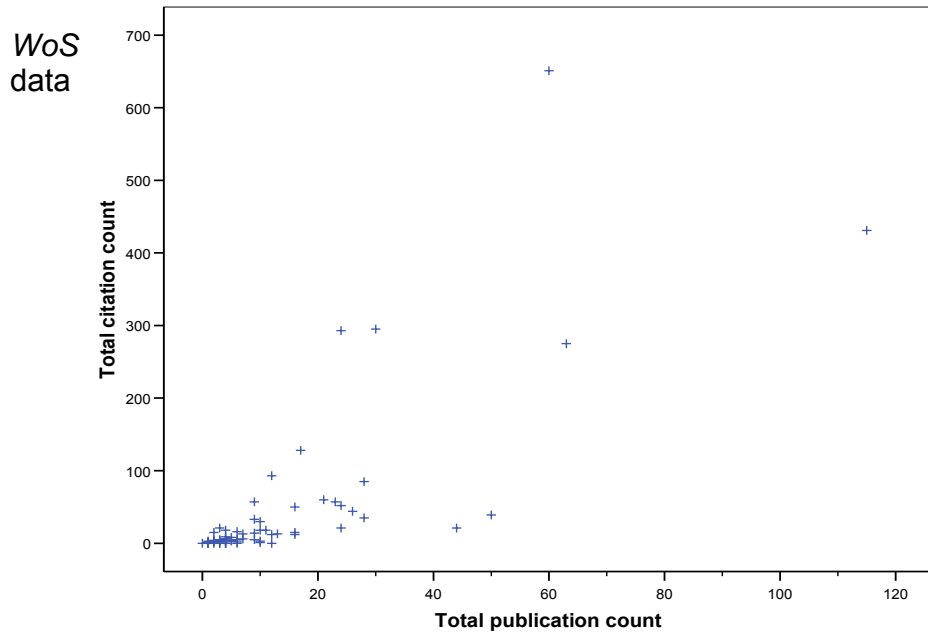
Figure 27. Scatter plots of total citation count against total publication count

Figure 27 confirms that total citation and total publication counts were positively associated.

4.3 Gender comparisons

4.3.1 Comparisons of metric scores for researchers of different gender
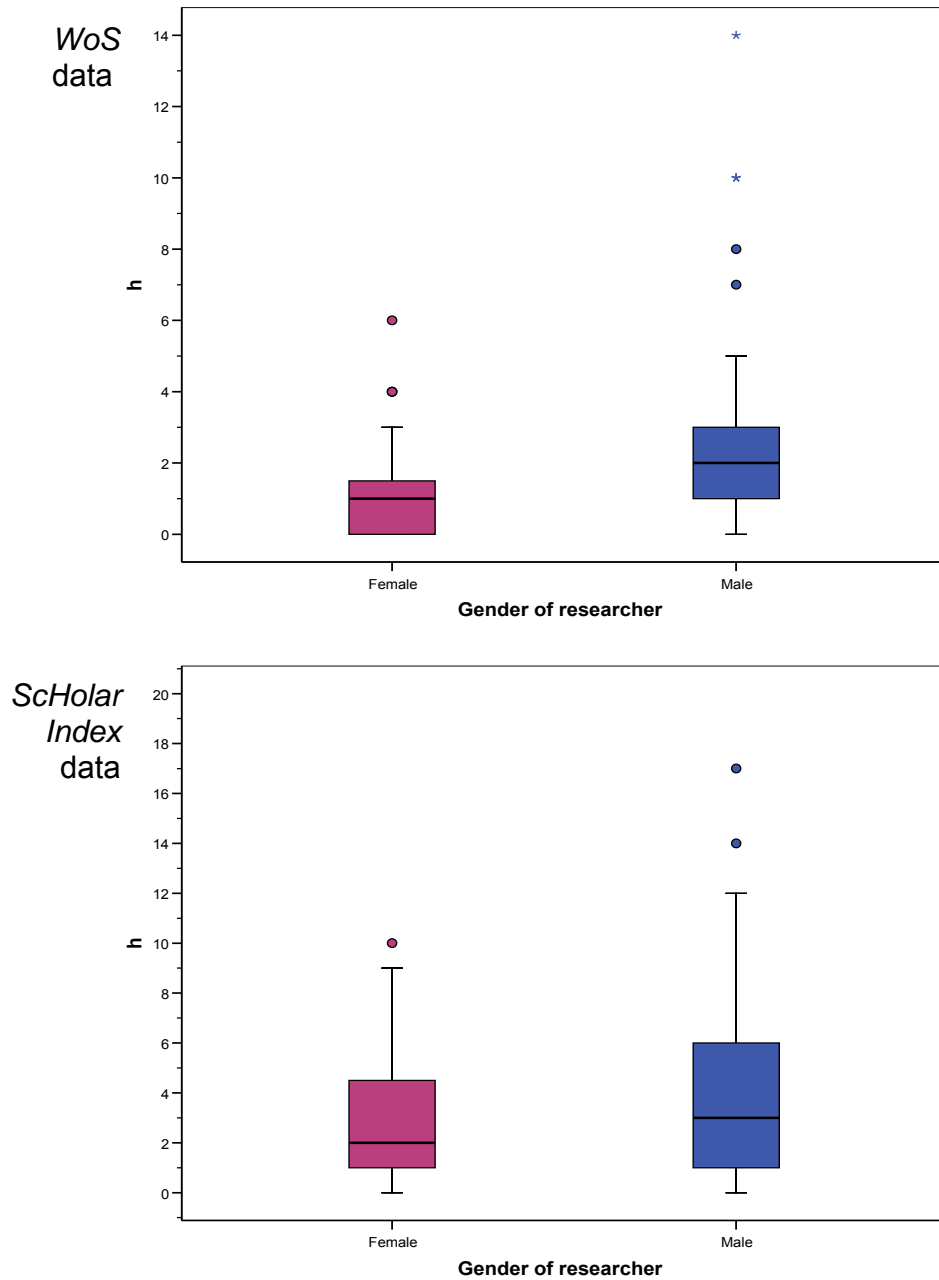
Figure 28. Comparative box plots showing median, inter-quartile ranges and ranges for *h* for different genders
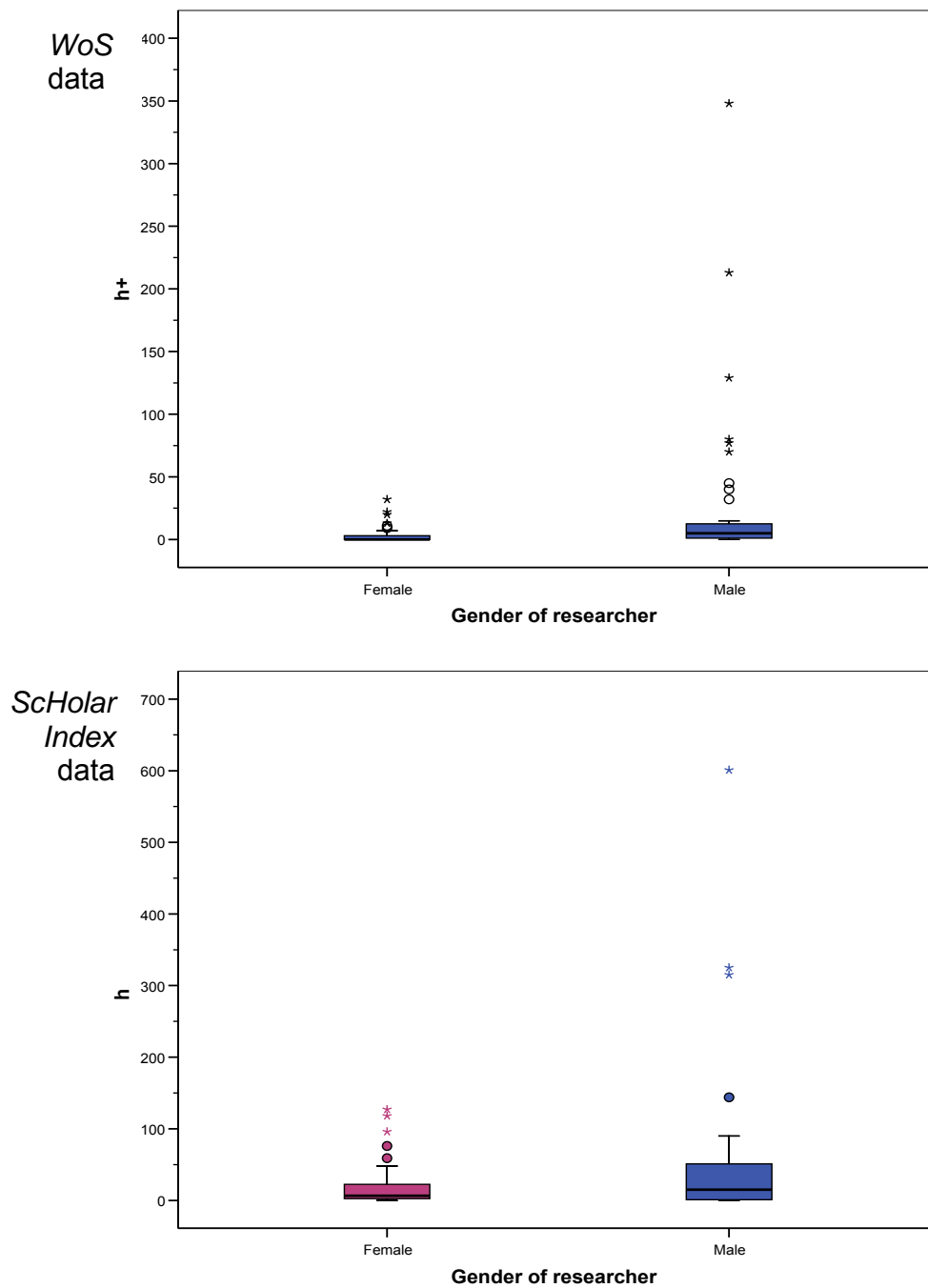
Figure 29. Comparative box plots showing median, inter-quartile ranges and ranges for h^+ for different genders
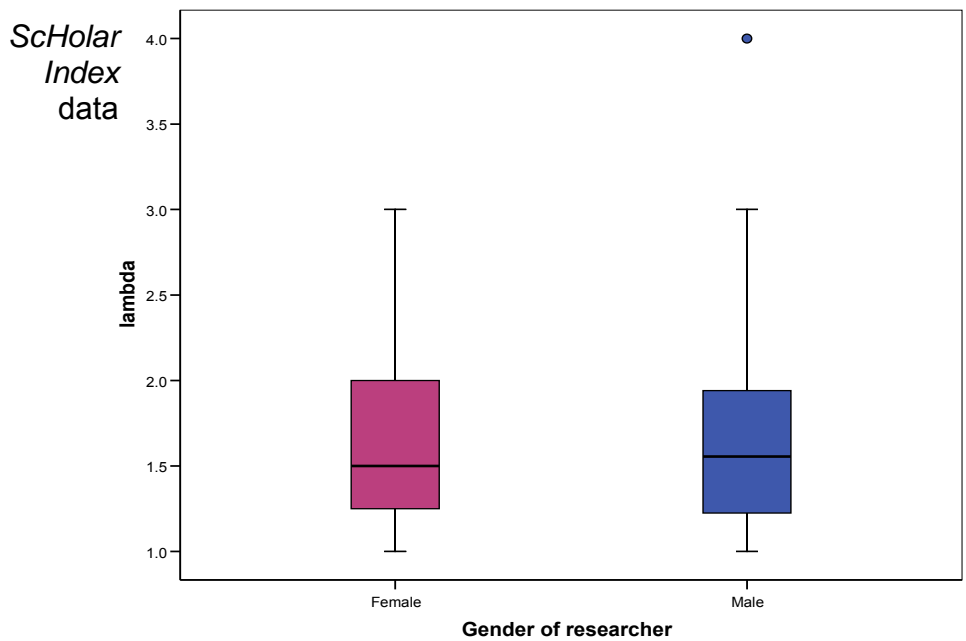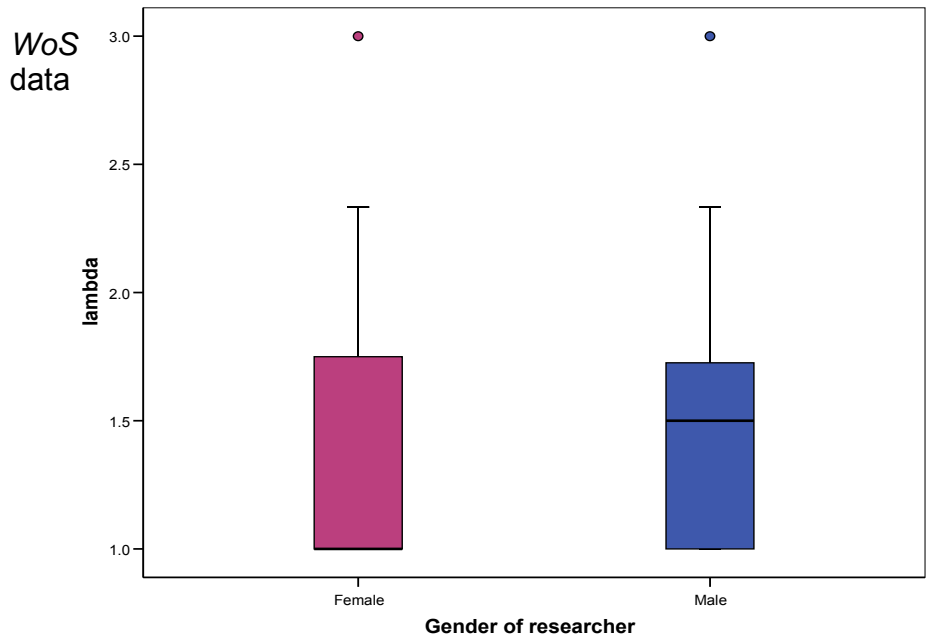
Figure 30. Comparative box plots showing median, inter-quartile ranges and ranges for Λ for different genders
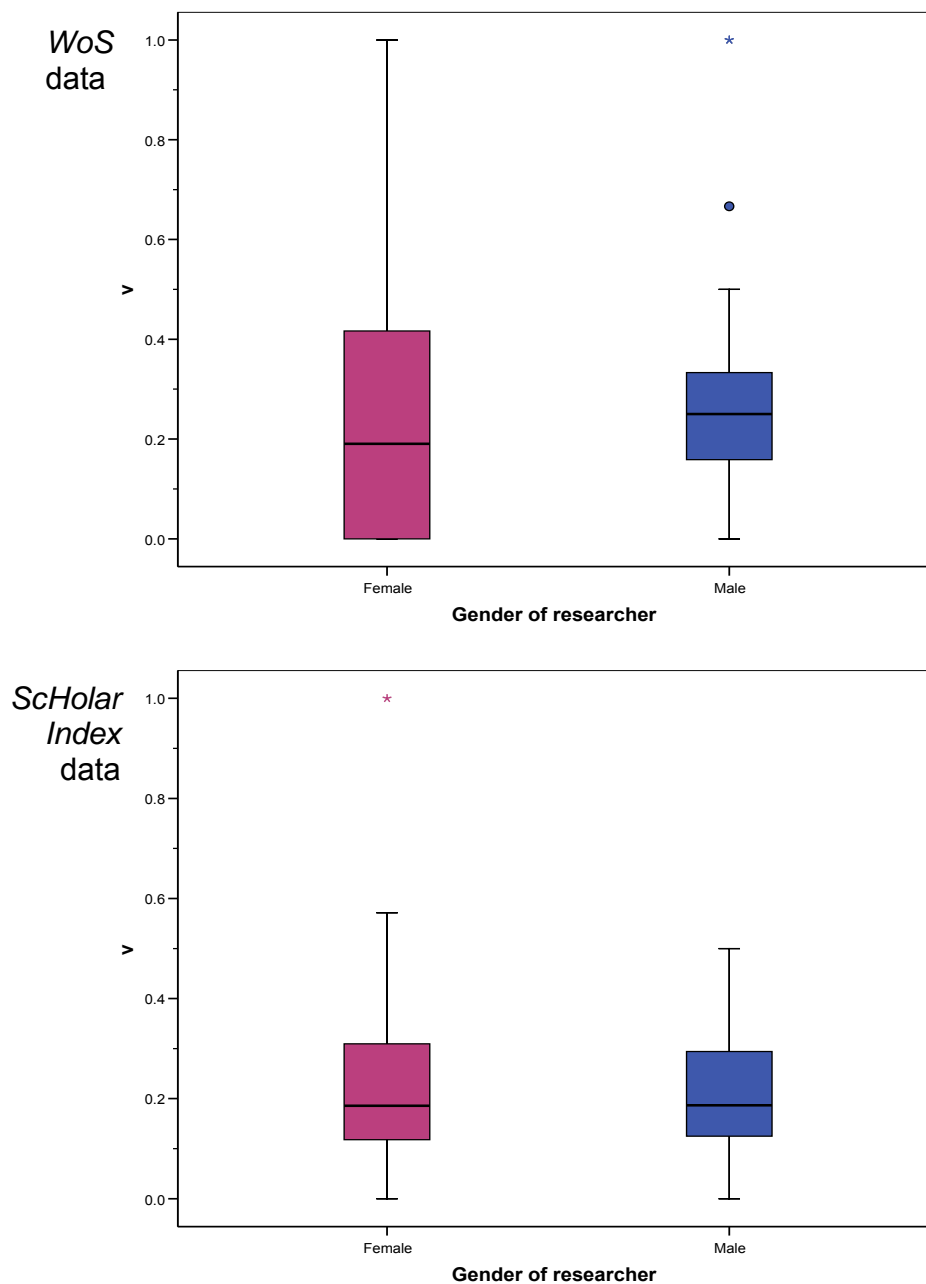
Figure 31. Comparative box plots showing median, inter-quartile ranges and ranges for *v* for different genders

Figures 28-31 show that the inter-quartile ranges for h, h^+, Λ and v scores were similar for male and female researchers, respectively, although the relative position of the medians varied. Most obviously, the median Λ for female researchers was far less than that for male researchers when calculated using *WoS*.

4.3.2 Comparisons of metric scores for researchers of different gender and academic rank
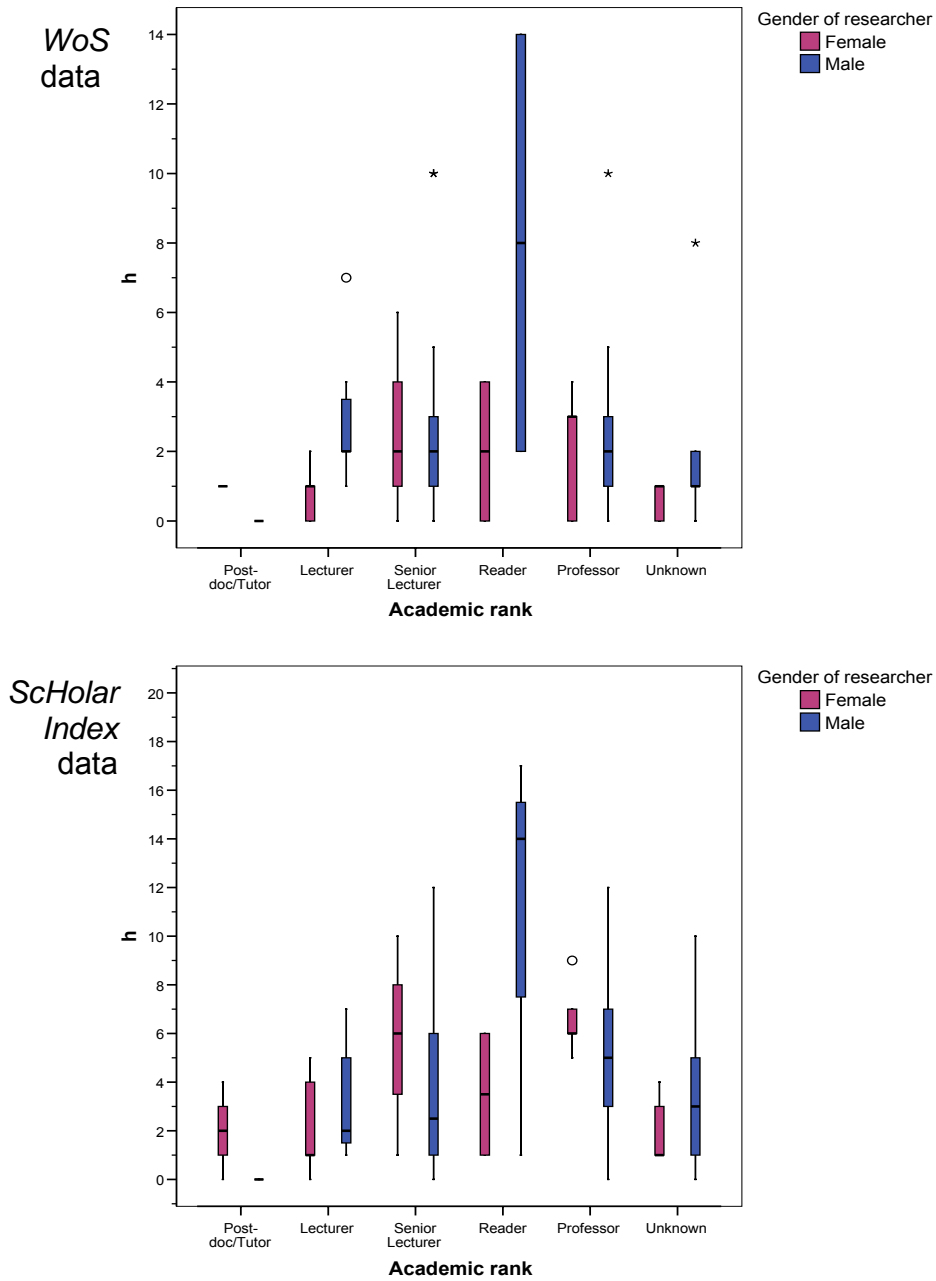
Figure 32. Comparative box plots showing median, inter-quartile ranges and ranges of *h* for different genders and academic ranks

Figure 32 shows that *h* varied with the database supplying the data. Male and female researchers shared similar *h*-index scores at all academic ranks except for Reader. Male Readers tended to have higher *h*-index scores than female Readers.
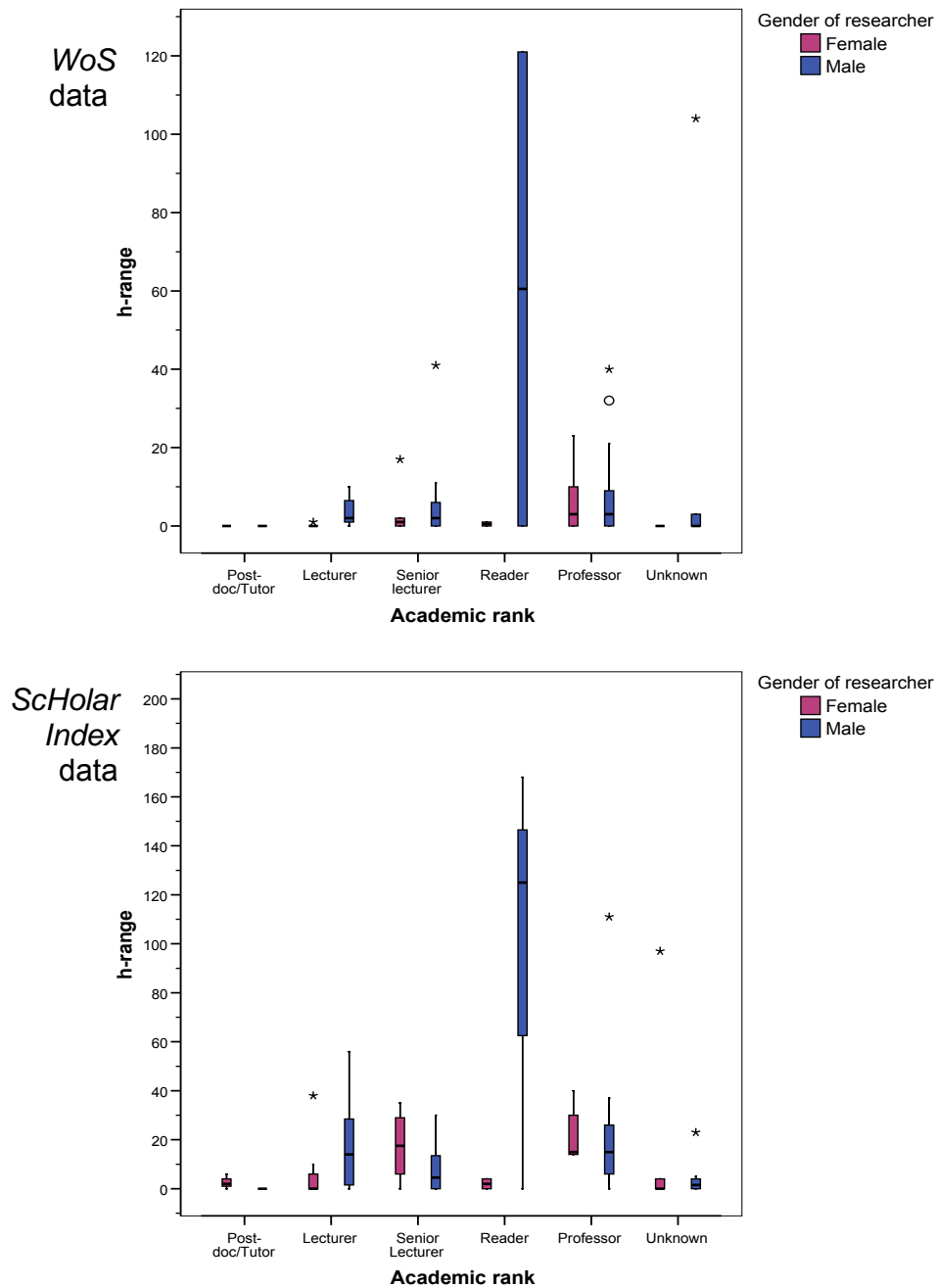
Figure 33. Comparative box plots showing median, inter-quartile ranges and ranges for *h*-ranges for researchers of different gender and academic rank
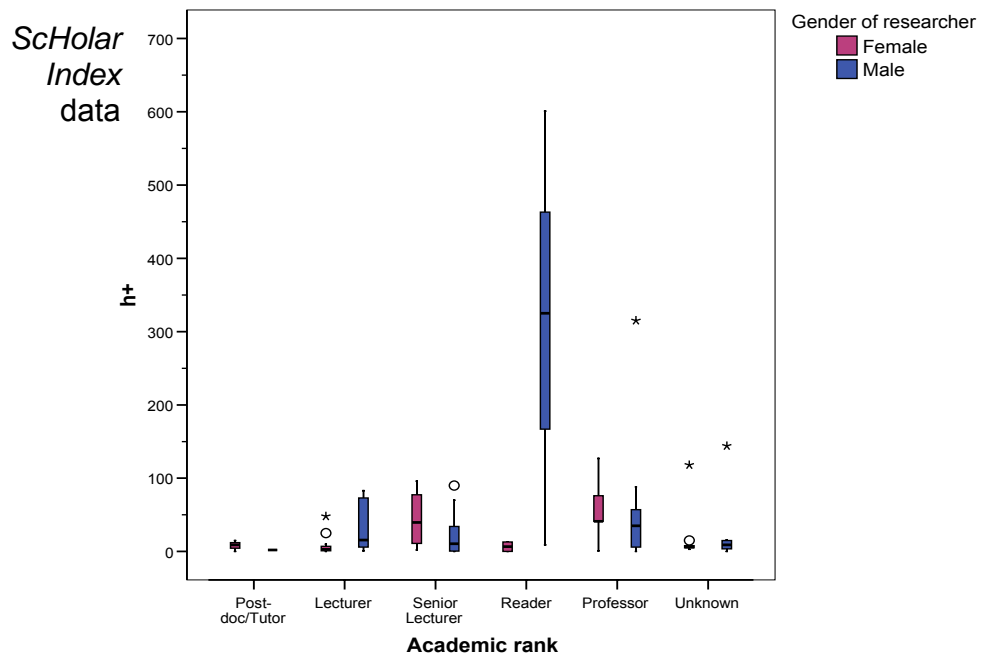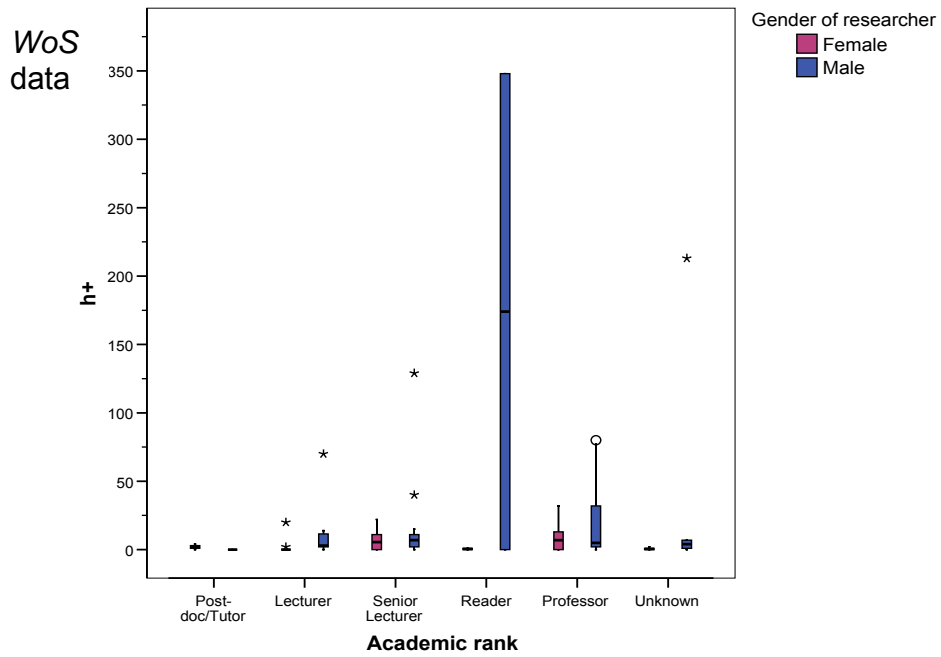
Figure 34. Comparative box plots showing median, inter-quartile ranges and ranges for h^+ for researchers of different gender and academic rank

Figures 33 and 34 show a similar pattern for h-range and h^+ as for h.
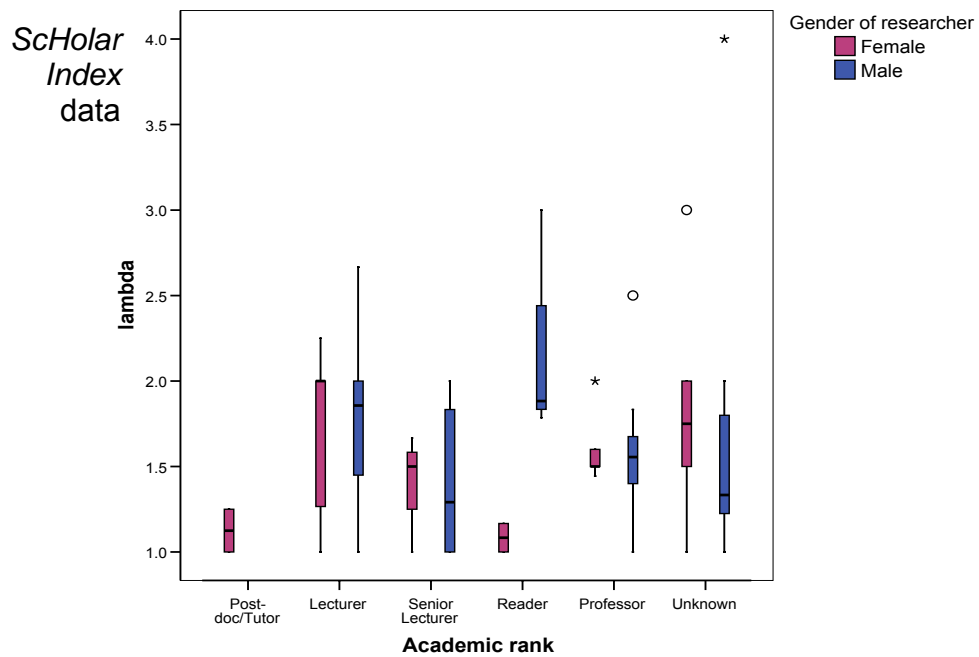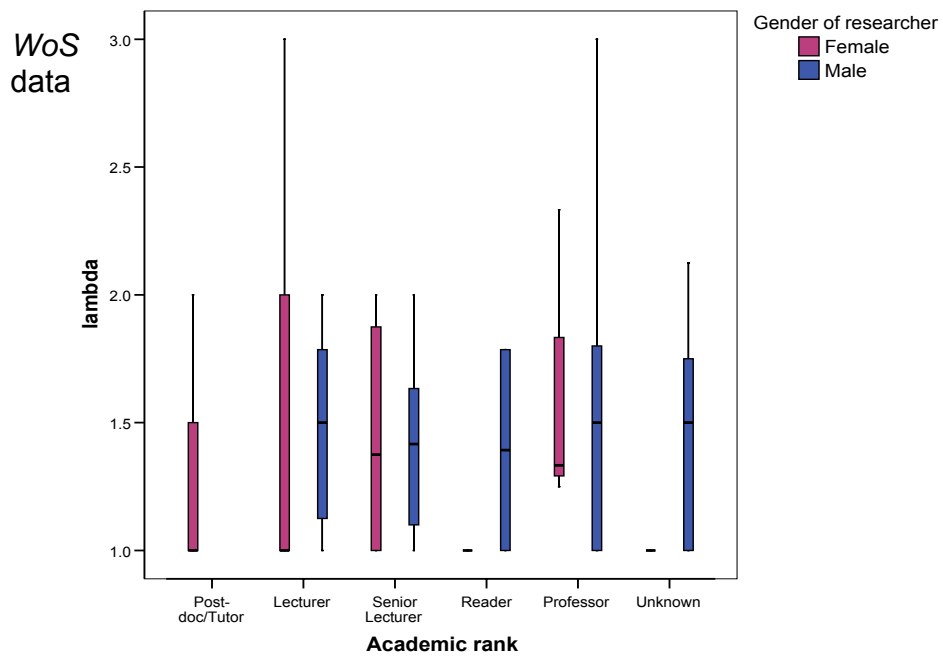
Figure 35. Comparative box plots showing median, inter-quartile ranges and ranges for Λ scores for researchers of different gender and academic rank

Figure 35 shows a similar pattern for Λ as for *h* when *WoS* data was used but the *ScHolar Index* data showed no difference in Λ between different genders or ranks.
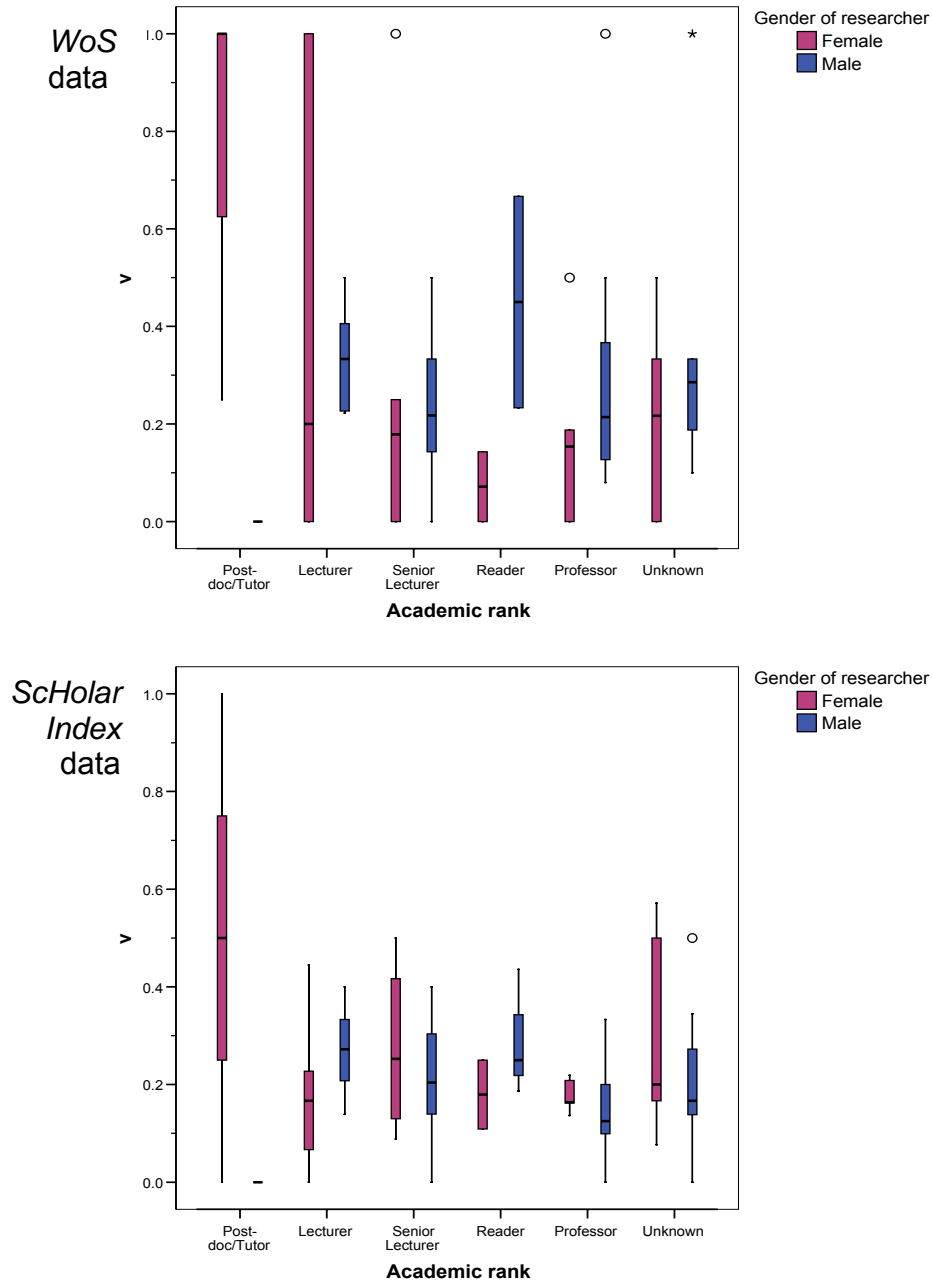
Figure 36. Comparative box plots showing median, inter-quartile ranges and ranges for *v* for researchers of different gender and academic rank

Figure 36 shows a similar pattern for *v* as for *h*.

4.3.3 Comparison of academic rank of different researchers of different gender at institutions of different RAE grade

Different numbers of female and male researchers were included in the analysis. It was therefore impossible to correlate random pairs at each institution. Instead, the ratio of female:male mean academic ranks of researchers at each institution were correlated against the RAE (2001) scores of their employing departments.
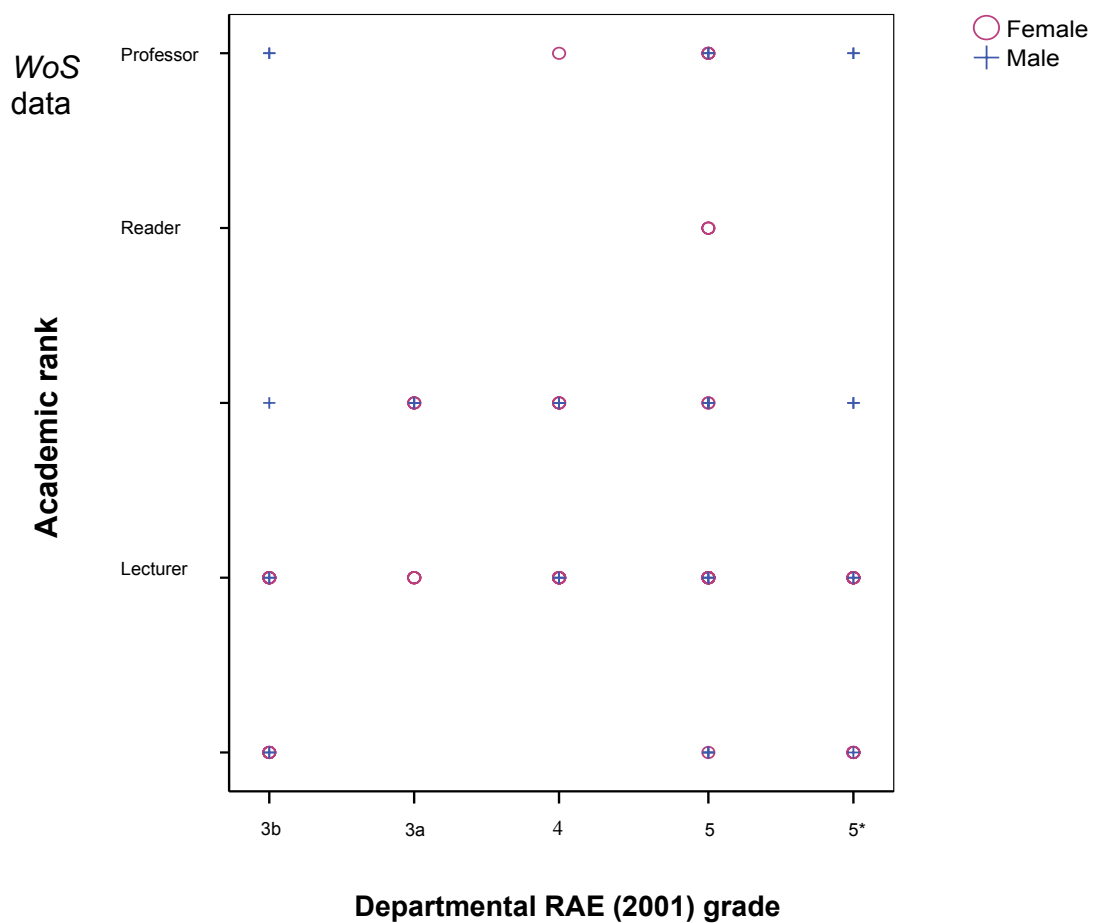
Figure 37. Scatter plot of academic ranks of all individuals of known academic rank against RAE (2001) grades of employing department
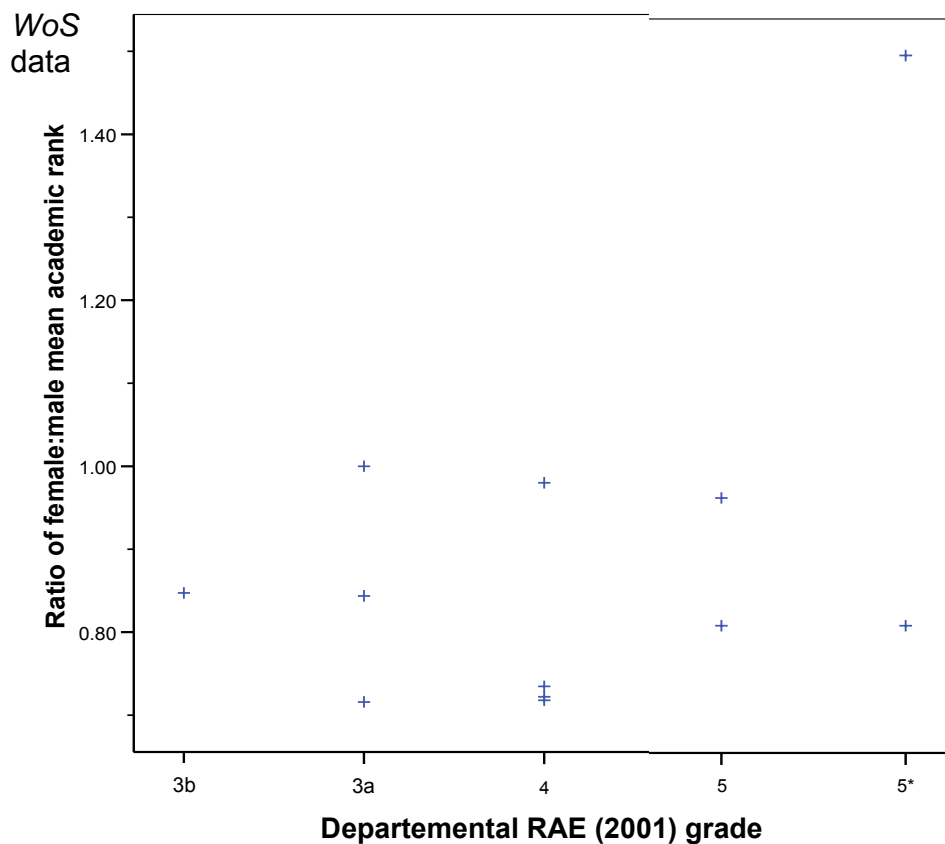
Figure 38. Scatter plot of all academic ranks of all individuals of known academic ranks against RAE (2001) grade of employing department

Figures 37 and 38 both demonstrate that absolute and relative academic advancement of female and male researchers is independent of the RAE score awarded to their employing departments. The ratio of female:male mean academic ranks of the 210 individuals of known academic rank was positively, although not significantly, correlated with departmental RAE score (r_S=0.647; P=0.148).

Individual institutions had a large effect on the overall score. When Napier University was excluded, the correlation coefficient became close to zero (r_S=-0.087; P=0.799 if the University of Wales, Aberystwyth was included and r_S=0.009; P=0.979 if only the English Universities were considered). The only Scottish

University for which academic positions were known (Napier University) had so many more highly ranking female researchers compared to male researchers than the other universities that it distorted the correlation statistic.
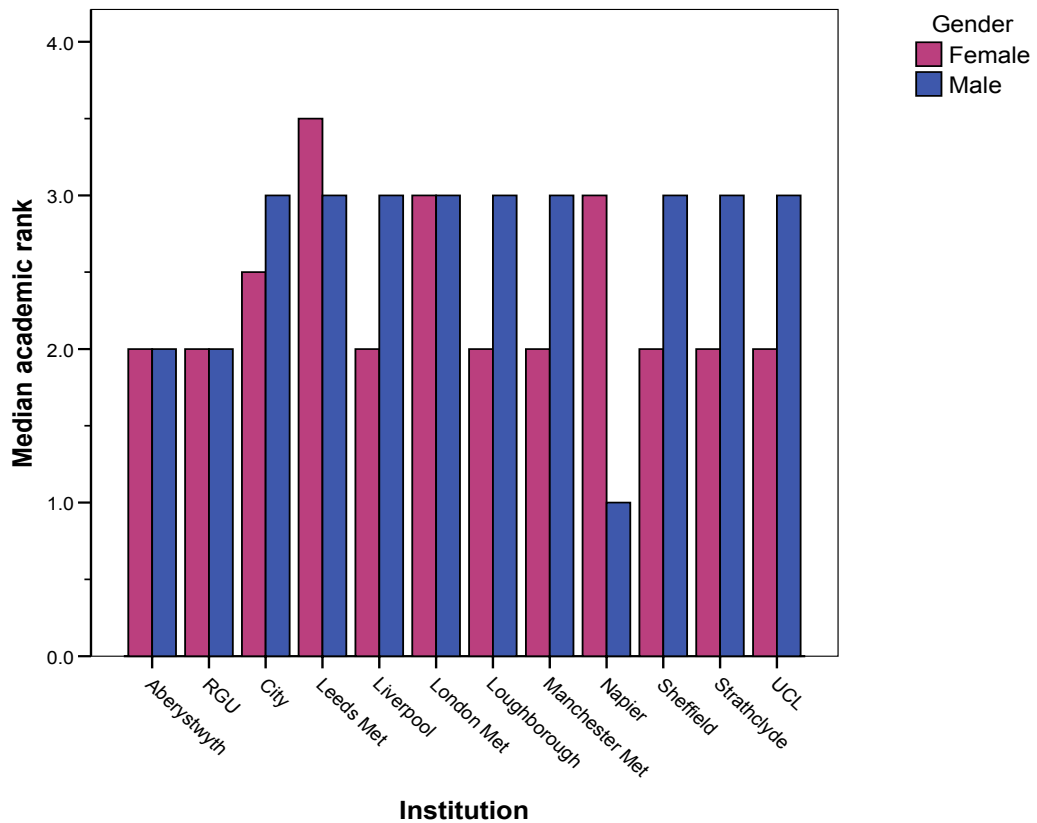
Figure 39. Comparative bar chart showing median academic rank of female and male researchers at different institutions.

Only one male researcher was employed by London Metropolitan University. His rank therefore constituted the median in Figure 39. At seven of the twelve institutions, the median male rank was one band higher than the median female rank. The median female rank was higher only at two institutions.

Friedman's test confirmed that the randomly selected women were ranked only slightly lower than men (median female academic rank=2.2500; median male academic rank=2.7500) and that the difference was not significant (S=1.00; P=0.317, adjusted for ties).

4.4 Comparison of average metrics at institutions of different RAE (2001) grade

4.4.1 Interaction between gender and departmental RAE (2001) grade
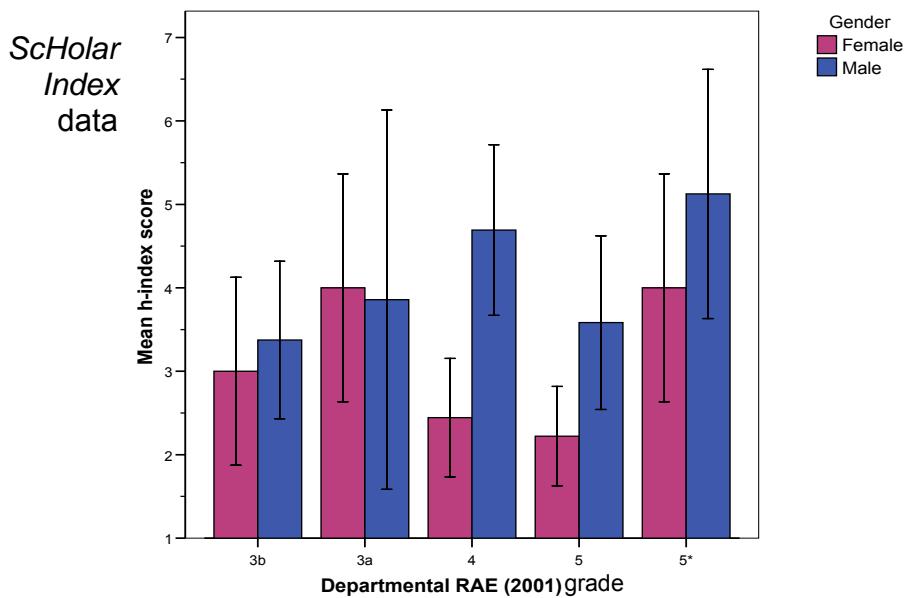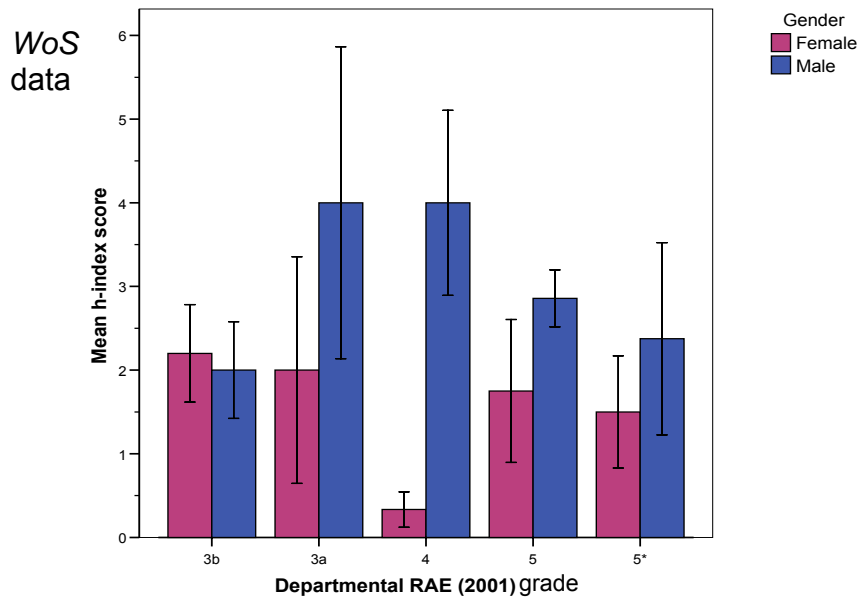
Figure 40. Comparative bar chart showing mean *h*-index scores (±1 S.E. of the mean) for different genders of researcher employed by institutions of different RAE (2001) grade
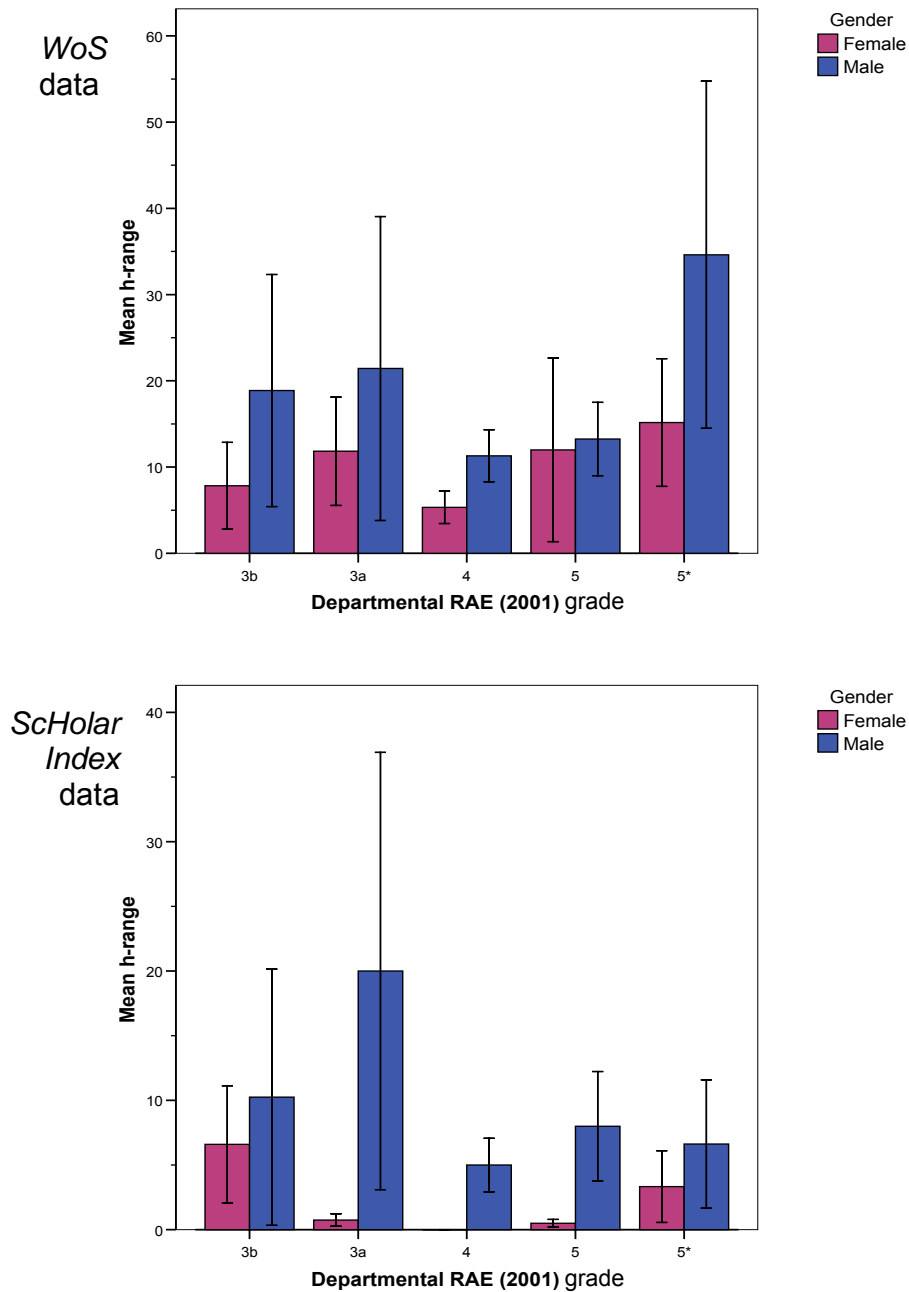
Figure 41. Comparative bar chart showing mean *h*-ranges (±1 S.E. of the mean) for different genders of researchers employed by institutions of different RAE (2001) grade
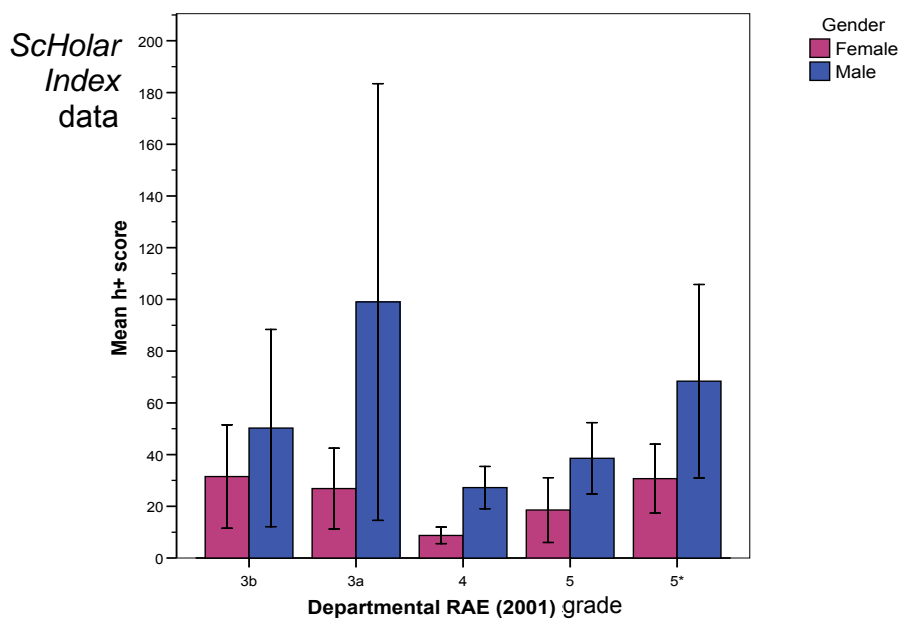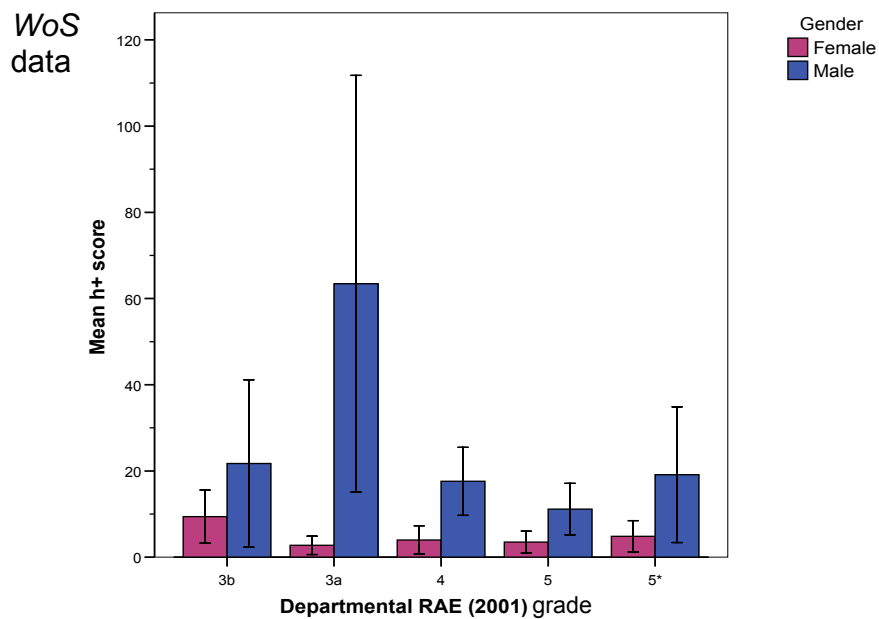
Figure 42. Comparative bar chart showing mean h^+ scores (±1 S.E. of the mean) for different genders of researchers employed by institutions of different RAE (2001) grade

Figures 40-42 show that mean h, h-range and h^+ scores for both male and female researchers varied with the database used.
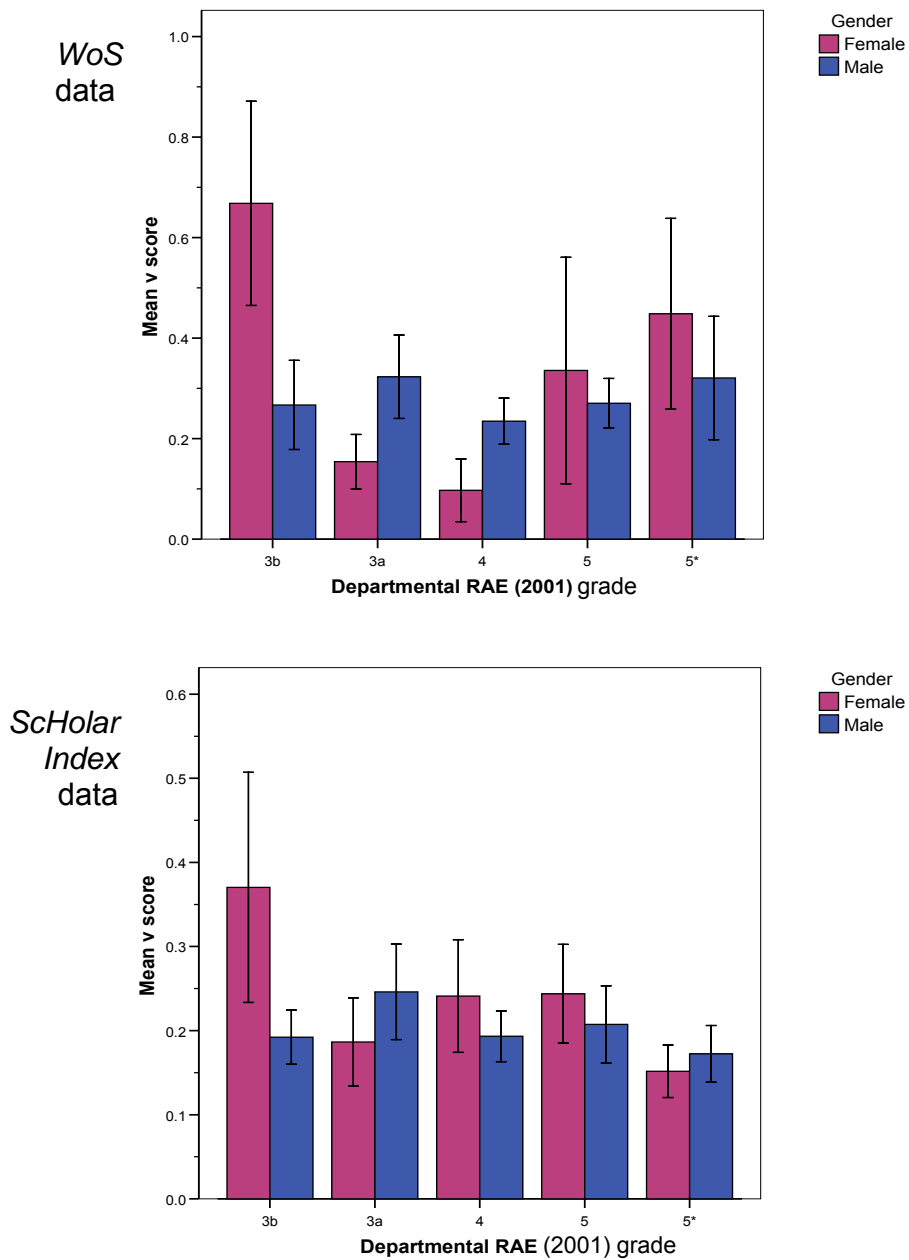
Figure 43. Comparative bar chart showing mean *v* scores (±1 S.E. of the mean) for different genders of researchers employed by institutions with different RAE (2001) grades

Figure 43 shows that in departments graded "3b", female researchers had slightly higher mean *v* scores than male researchers but at all other institutions, there appeared to be little difference between genders.

WoS data

ScHolar Index data
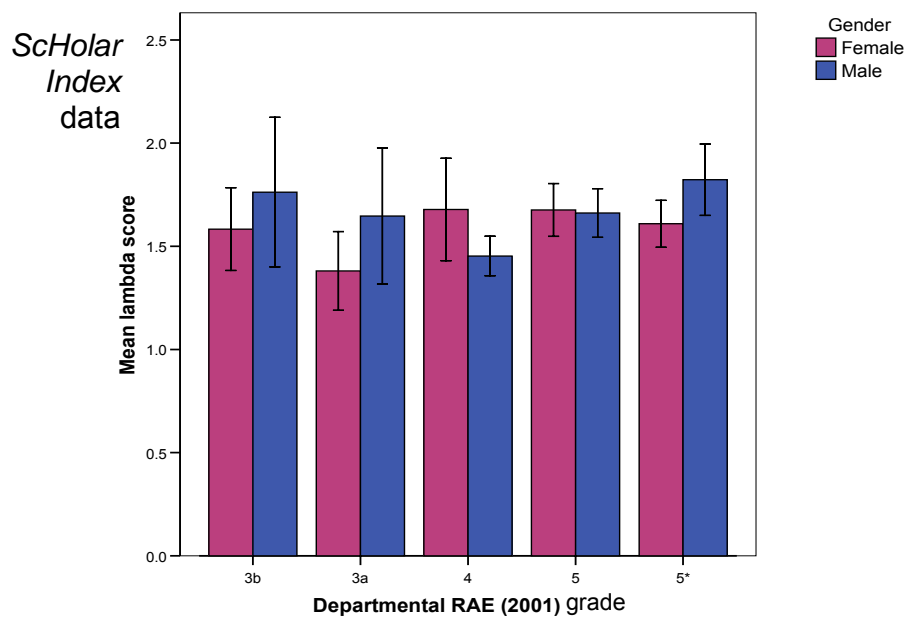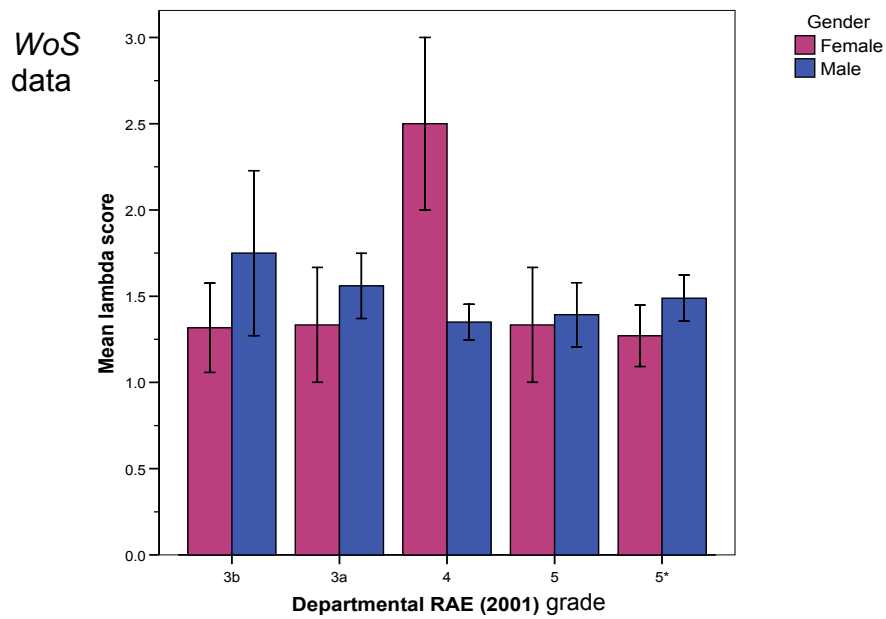
Figure 44. Comparative bar chart showing mean Λ scores (±1 S.E. of the mean) for different genders of researchers employed by institutions of different RAE (2001) grade

Λ varied very little between institutions (Figure 44) and, despite disagreement between the databases about the mean female Λ score at institutions graded "4" for research, there was little difference between the genders. The standard errors for the metrics were large and overlapped for male and female researchers, suggesting that the mean metrics for both genders were the same. A Sheirer-Ray-Hare test showed that, for the *Web of Science* results, only h^+ scores varied significantly and only between departments of different RAE grade ($\chi^2_{v=4}$=0.970; P<0.030). The *ScHolar Index* results showed no significant difference between the metric scores of different genders or departments of different RAE grade (Appendix H).

4.4.2 Correlation of mean metrics against RAE (2001) grade of employing department

WoS data

r_S = 0.211; P = 0.451
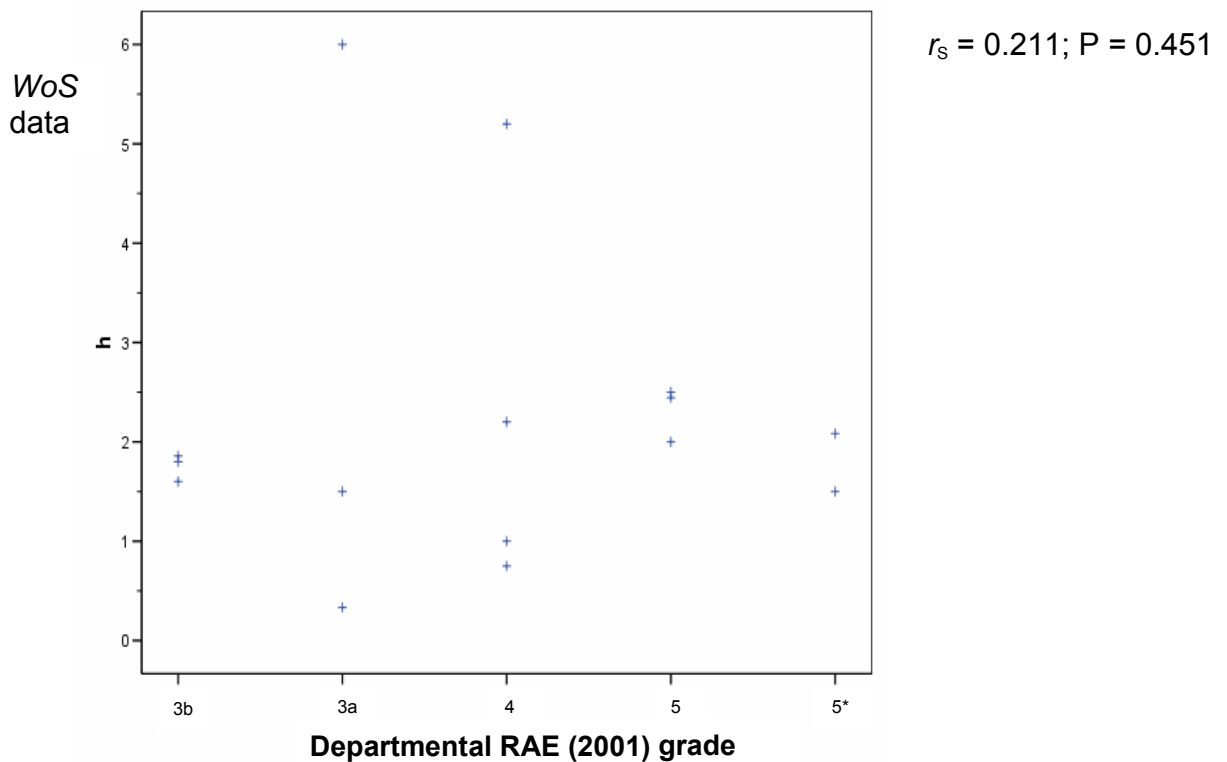
Figure 45. Scatter plot of *h*-index scores of researchers against RAE (2001) grade of employing department
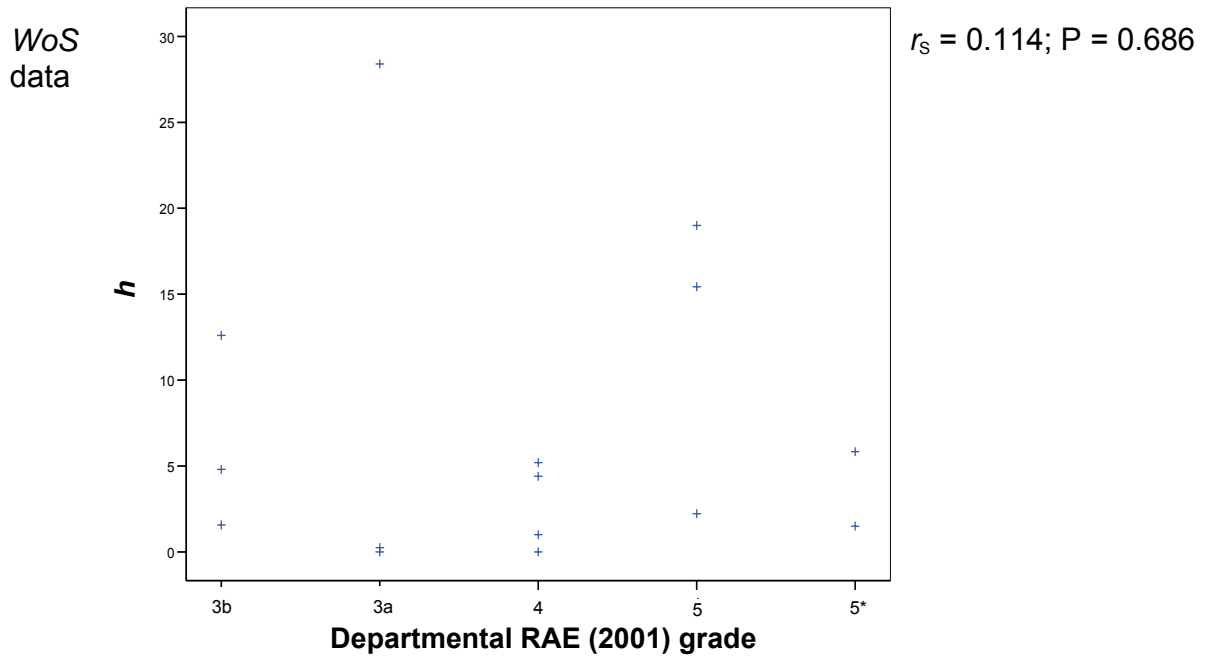
Figure 46. Scatter plot of *h*-range scores of researchers against RAE (2001)grade of employing department
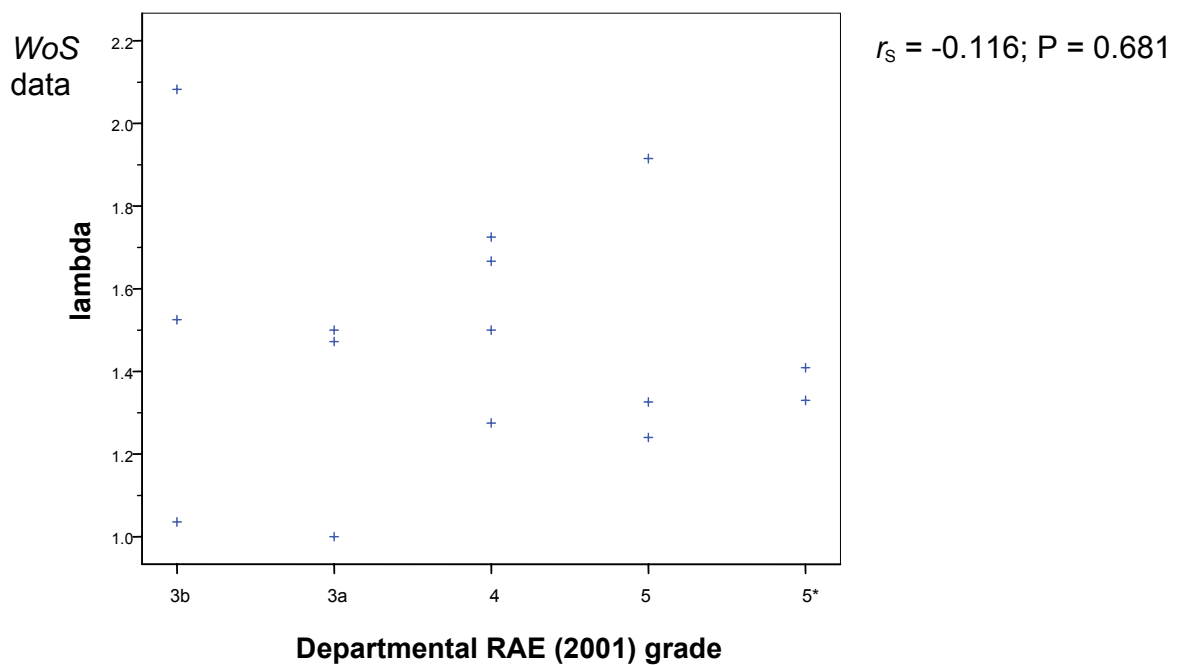
Figure 47. Scatter plot of Λ scores of researchers against RAE (2001) grade of employing department
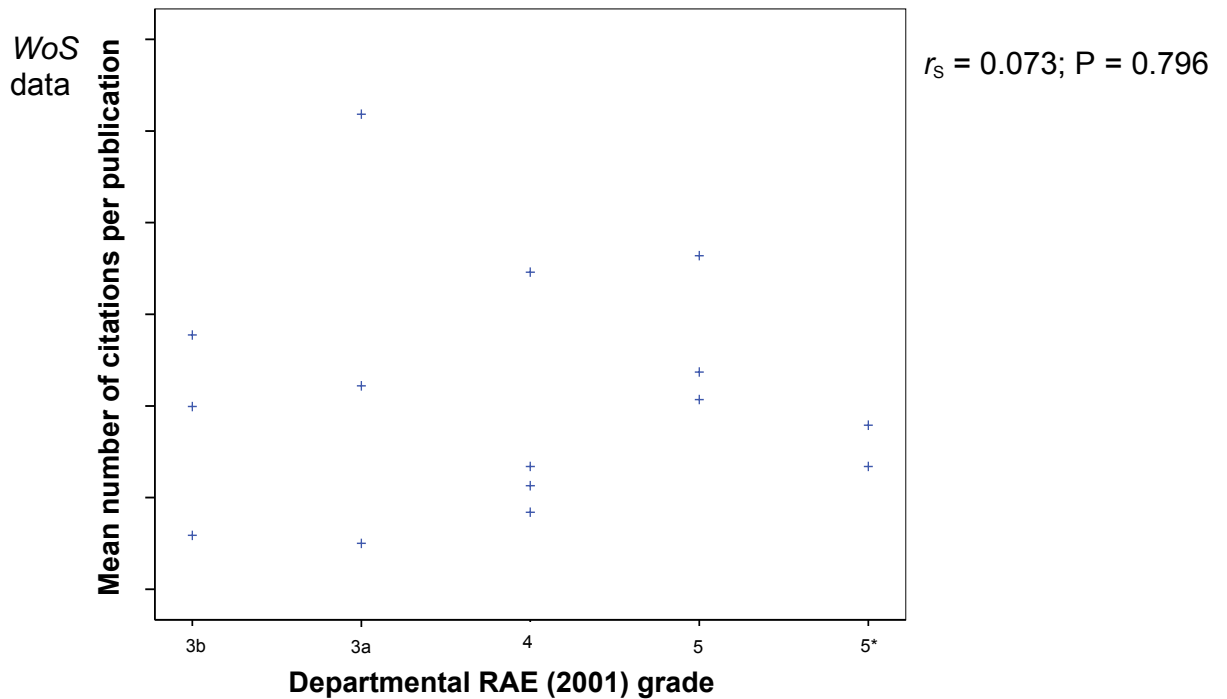
Figure 48. Scatter plot of the mean number of citations per publication against RAE (2001) grade of employing department

Figures 45-48 confirmed that there was no correlation between the RAE (2001) grade of the employing department and h, h-range, Λ, or mean number of citations per publication. Neither mean departmental total citation counts (r_{S}=0.303; P=0.136) nor total publication counts (r_{S}=0.257; P=0.177) correlated with RAE (2001) grade (Appendix D).

4.5 Comparison of peer and *h*-index rankings

Table 3 shows that the peer rankings of thirteen individuals were positively correlated with corresponding *h*-index based rankings ($r_S \geq +0.3$). Although nine of these were quite strongly correlated ($r_S \geq +0.6$), only three were significantly correlated.

There appeared to be no overall agreement between peer and *h*-based ranking data, overall or for researchers of different academic ranks. Figure 49 confirmed no association between academic rank and agreement of peer and *h* based ranking of researchers. Senior Lecturers disagreed most with the *h*-index based rankings. A Kruskall-Wallace test showed that the median correlation coefficients did not vary significantly between individuals of different academic rank ($\chi^2_{v=4}$=0.953; P=0.917).
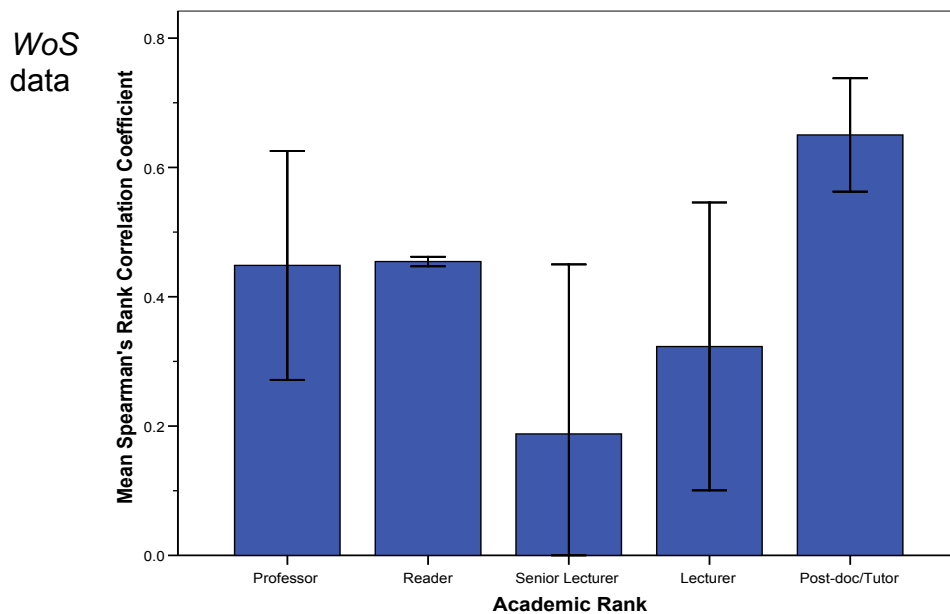
Figure 49. Mean r_S values with standard errors of correlations of peer-ranked and *h* ranked individuals for respondents of different academic rank

Table 3. Frequencies of Spearman's Rank correlation coefficients for *h*-index and peer rankings of researchers, broken down by approximate academic rank of the researcher providing the peer rankings

Rank of individual ranking their peers	Frequencies of correlations with Spearman's rank coefficients of different sign and magnitude					
	$-1 \leq r_S < -0.3$	$-0.3 \leq r_S < +0.3$	$+0.3 \leq r_S < +0.6$	$+0.6 \leq r_S \leq +1$	Of which significant	Total
Professor	0	2	1	2	1	6
Reader	0	0	2	0	0	2
Senior Lecturer	0	1	1	2	0	4
Lecturer	1	2	0	3	1	7
Post-doc/Tutor	0	1	0	2	1	4
Total	1	6	4	9	3	23

* *5 per cent significance level (1-tailed)*

4.6 Gender analysis of peer rankings

Table 4 shows that the relative frequency of different gender researchers in the top-five rankings for each department was almost exactly proportional to the number of researchers of each gender in the sample frame (Table 4). This was confirmed by a chi-squared goodness-of-fit test $(\chi^2_{v=1}=0.0000469$ (3 s.f.); P>0.050). This comparison was not entirely accurate because some peer-ranked data included individuals from their departments that were outside of the sample frame.

Table 4. Frequencies of male and female researchers who were included by peers in the top five researchers in their department (peer rankings) and frequencies of male and female researchers of senior lecturer rank or above in the sample frame

Data source	Peer rankings (top five researchers in department)		Sample frame	
Gender	Male	Female	Male	Female
Frequency	74	29	23	9
Percentage (2 d.p.)	71.84	28.16	71.88	28.13

4.7 Correlation of paired metric scores calculated using different citation databases
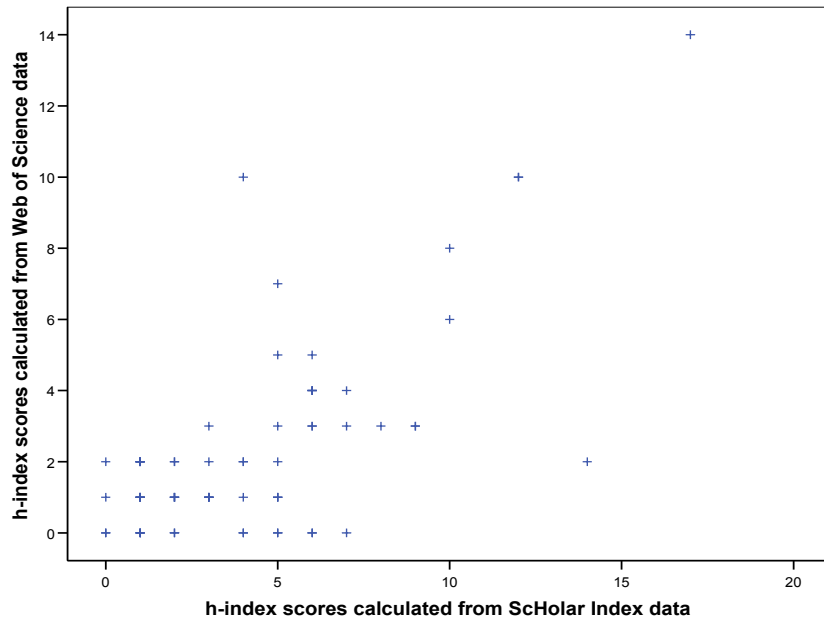
Figure 50. Scatter plot of *h*-index scores generated from *WoS* and *ScHolar Index* data
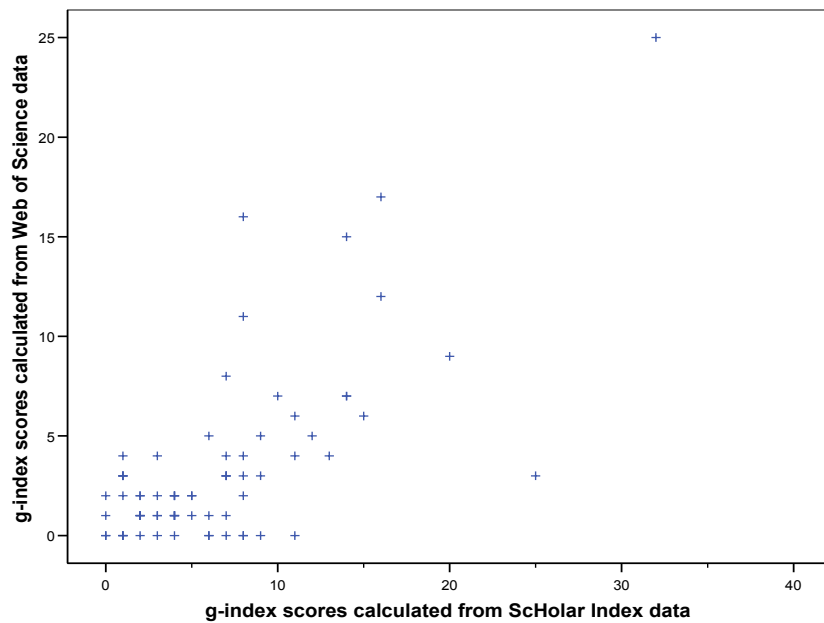
Figure 51. Scatter plot of *g*-index scores generated from *WoS* and *ScHolar Index* data
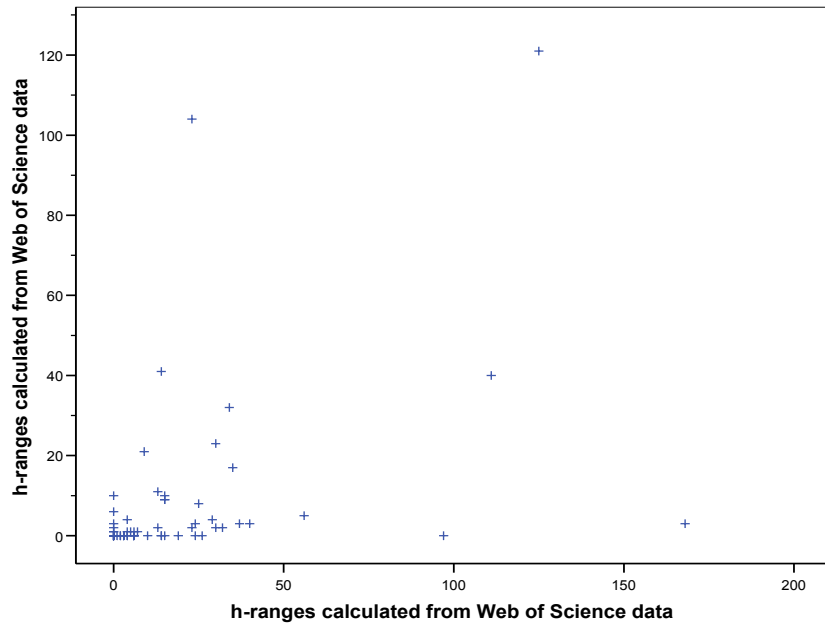
Figure 52. Scatter plot of *h*-ranges generated from *WoS* and *ScHolar Index* data

Figures 50-52 clearly show that the *h*, *g* and *h*-ranges calculated using data from the two citation databases were strongly correlated but that much of the data was clustered near the origin.
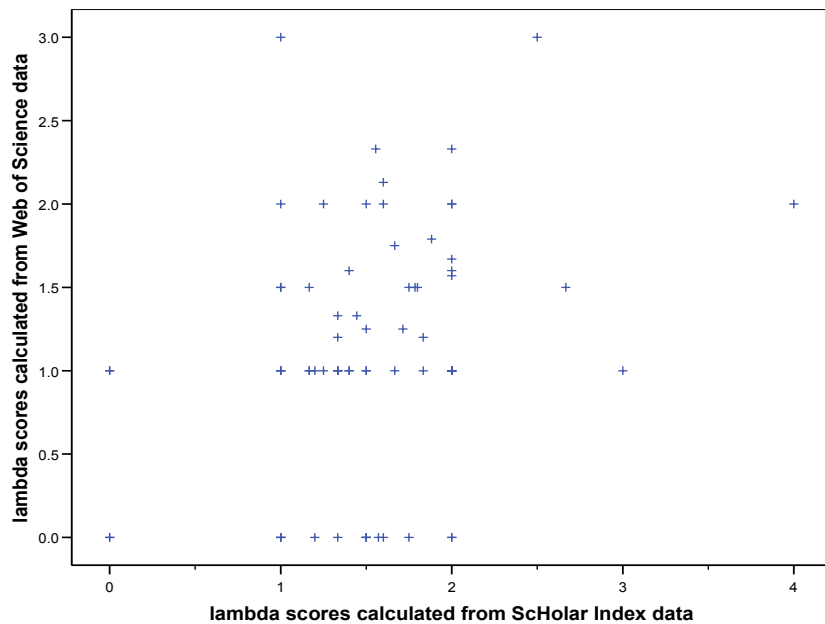
Figure 53. Scatter plot of Λ scores generated from *WoS* and *ScHolar Index* data
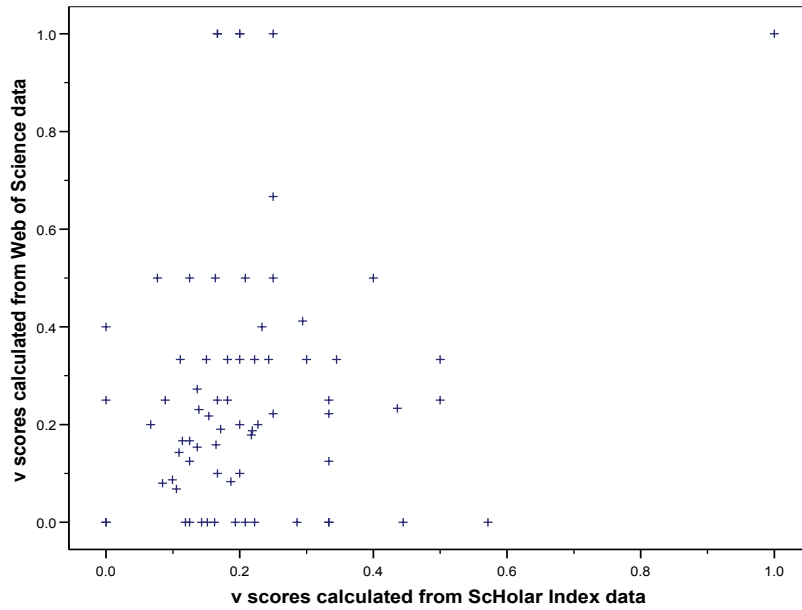
Figure 54. Scatter plot of *v* generated from *WoS* and *ScHolar Index* data

Figures 53-54 show that the Λ and *v* scores calculated using the two databases were unpredictably associated, despite being significantly correlated (r_S=0.273; P=0.021 and r_S=0.172; P=0.153, respectively).
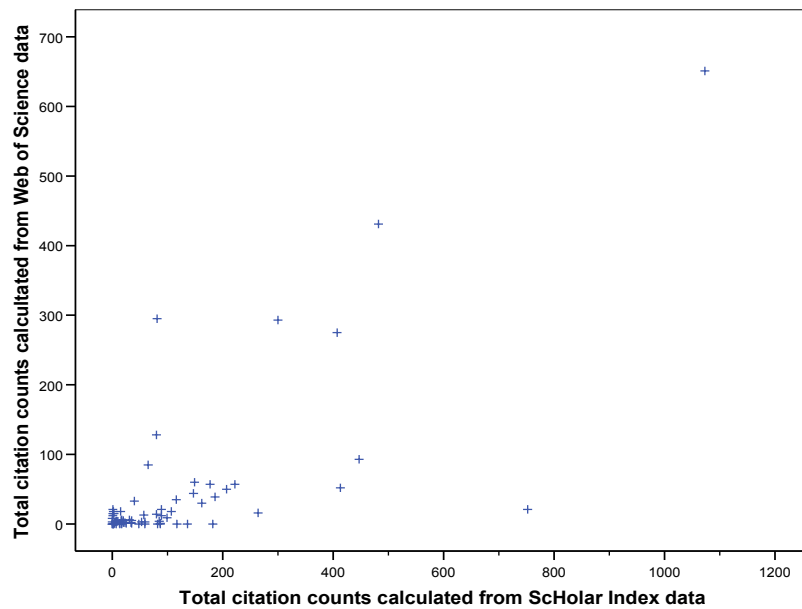
Figure 55. Scatter plot of total citation counts generated from *WoS* and *ScHolar Index* data
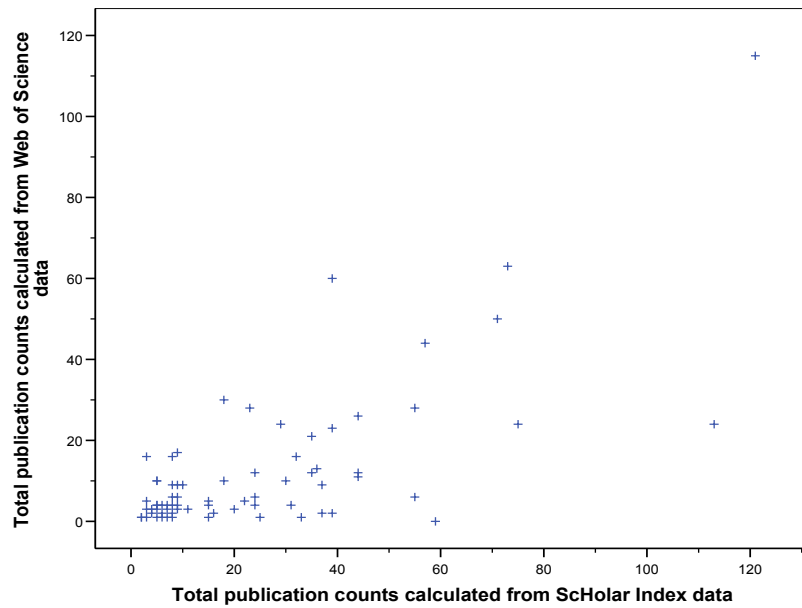
Figure 56. Scatter plot of total publication counts generated from *WoS* and *ScHolar Index* data

Figures 55-56 show that both total citation and publication counts positively correlated between the two databases.
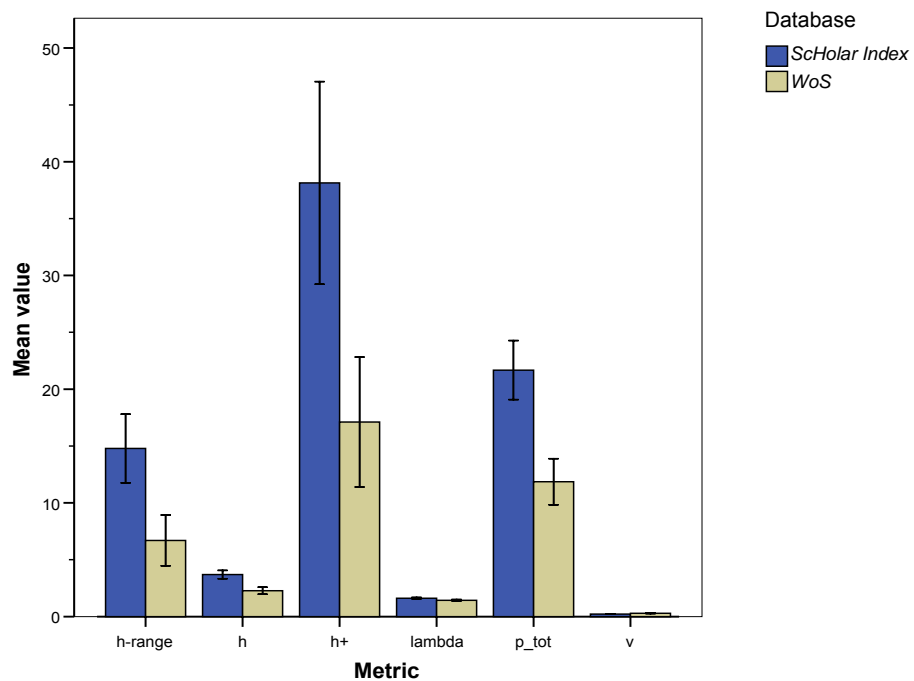
Figure 57. Comparative bar chart showing values for different metrics calculated from *WoS* and *ScHolar Index* data. Bar heights represent mean metrics; lines represent one standard error either side of the mean

Figures 57 and 58 show that the metrics calculated using *ScHolar Index* were, on average, greater than those calculated using *Web of Science*. The standard errors of the means of each metric calculated from data from the two databases did not overlap, suggesting that the underlying population means were probably different.

A Sheirer-Ray-Hare test confirmed that the median metrics calculated using the two databases were very highly significantly different (SS_{metric}/Total MS=73.819 (3 d.p.); P<0.001).
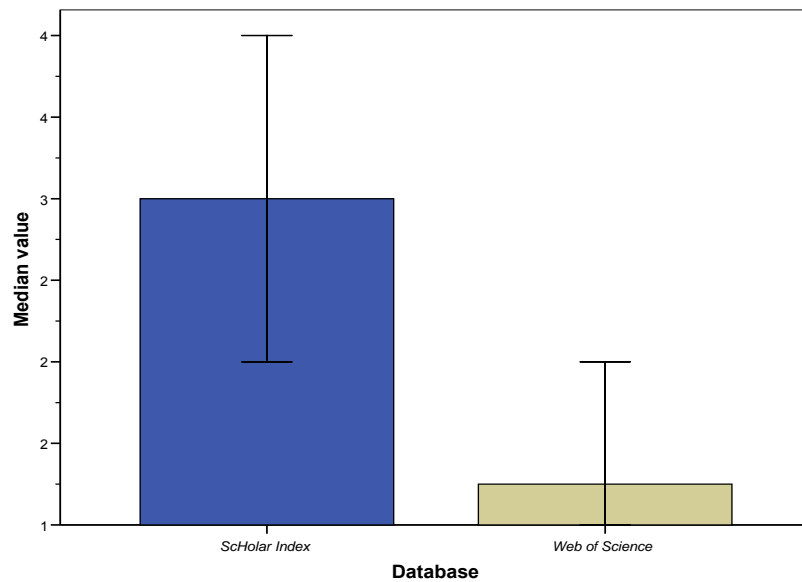
Figure 58. Bar chart showing medians ranges of *h, h*-range, *h⁺, v, Λ*, total citation count and total publication count metrics calculated from *Web of Science* and *ScHolar Index* data
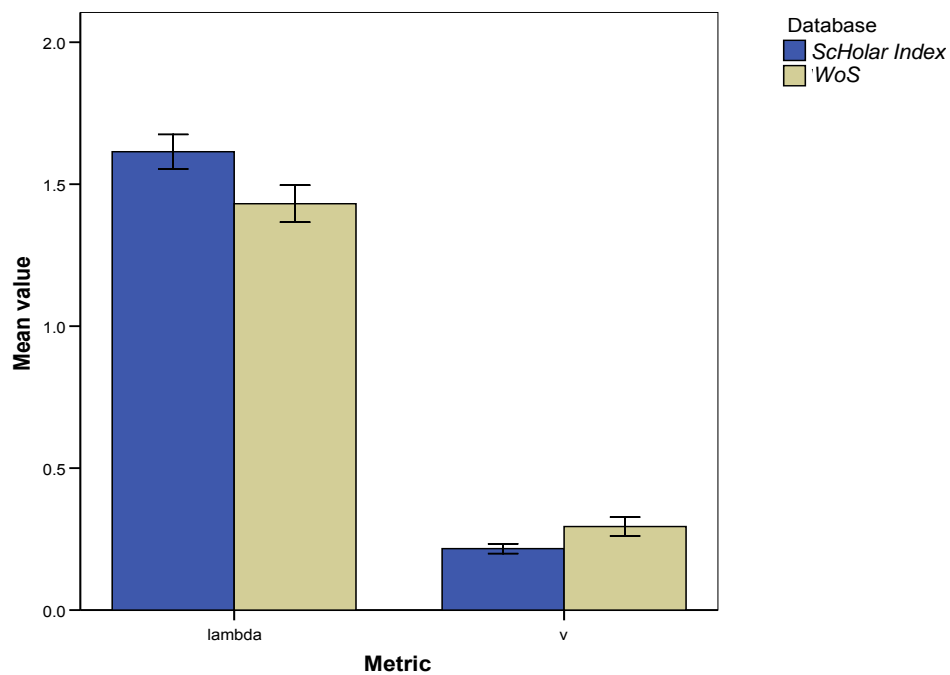
Figure 59. Comparative bar chart showing Λ and *v* calculated using *WoS* and *ScHolar Index* data

Figure 59 shows that Λ and *v* scores generated using *Web of Science* were very slightly larger than those generated using *ScHolar Index* data, although the gap between the extremes is small compared to the standard errors of the means.

4.8 Qualitative analysis

Responses were read, coded using natural language and then categorised. Most categories, but not all, were saturated.

Tables 5 and 6 suggest that Library and Information Science researchers were well informed about citation analysis and were equally critical of both peer assessment and citation-based assessment for different reasons, leading a few researchers to suggest that they should both be used in tandem.

This was borne out by a fairly even split between those respondents who openly supported peer assessment over citation analysis and those who supported both methods in tandem (Table 8).

Table 5. Frequency of comments about peer assessment

Advantage	Frequency	Disadvantage	Frequency
Emotional support and professional guidance	2	Subjective assessment bias	5
Holistic appraisal method	2	Corrals research into popular fields and opposes challenges experimentation and challenges to orthodoxy	9
Discourages interdisciplinary work	1	Matthew effect (visibility and influence effects) evident	1
		Old Boy Network operates	4
		Subject to fashions	2
		Personal, political and inter-personal biases operate	12
		Includes 'soft' factors	1
		Expensive	2
		Skewed by superficial knowledge	1
		Lack of expertise	5
		Disagreement of author and reviewer prevents publication at peer review stage	2
		Visibility/influence bias favours established researchers	4

Table 6. Frequency of comments about citation analysis

Advantage	Frequency	Disadvantage	Frequency
Objective	3	Biased for reasons discussed in Chapter 2	14
Can indicate breadth of influence	1	Not a measure of research quality	1
		Corrals research into popular fields and opposes challenges experimentation and challenges to orthodoxy	9
		Matthew effect evident	3
		Visibility/influence bias	17*
		Favours interdisciplinary research	1**
		Negational citations	10
		Retrospective bias (favours older research)	4
		Overly heuristic	1
		Sub-discipline specific	3
		Measures productivity, not citedness in Humanities disciplines	2
		Vulnerable to manipulation	3
		Citation databases biased both for coverage and for English language publications	6
		Disagreement causes non-citation	1
		Groups and individuals cite one another reciprocally	5

* The identification of "key" researchers and works by citation analysis ignores recently published works, especially by younger researchers, and useful works of poor visibility.

** Any bias that favoured one form of research practice over another was assumed to be undesirable in a performance measure.

Many participants regarded peer review a major part of peer assessment.

Reasons given for favouring peer assessment included the private personalised feedback and support to improve an individual's research performance. Several individuals commented on the relatively inexpensive and rapid calculation of citation metrics as compared to peer assessment, although citation databases were criticised for the length of time they took to add citations, exacerbating the retrospective nature of citation metrics, which, like peer review, only recognise past performance and not potential. Others criticised citation metrics for exacerbating the "Matthew effect" and for assessing researchers on spurious grounds, although others argued that citation metrics and peer assessment both measure author visibility.

The consensus appeared to be that rather than measuring impact, citations measured "interest" in a publication, which would be a questionable basis for assessment, and might even be considered "subversive".

In addition to negational citations, the heuristic meaning of positive citations was also felt to make them unclear, particularly where reference was made to standard methods as described in another paper, although such citations still show intellectual heritage.

Influence effects were also held up as citation biases, with suggestions that students and supervisors cited one another unduly, although the publications were probably highly relevant and therefore valid to cite.

Both methods of assessment were criticised for suppressing creativity and enforcing norms, since the same individuals in a small subfield that assess work are also those who will choose whether or not to cite it, such that some conformity to orthodoxy is required in order to pass peer review or to become cited, respectively. Citation metrics therefore appear to offer a similar critical audience in small subfields as peer assessment. Criticism of peer assessment

mainly attacked the lack of knowledge and prejudice of publication reviewers than of general performance assessments.

The rigour of peer review, which is critical to minimising the number of purely critical negational citations and maintaining the clarity of meaning that citations should have, was also criticised. Individuals' publication counts and the stage of development of the field in which they work, *i.e.* their visibility affects how easy it is for them to get further research published and cited.

Some participants offered tentative conclusions, which are summarised in Table 7. Categories are not mutually exclusive.

Table 7. Participants appraisal of peer review and citation analysis

Comment	Frequency
Both methods are robust	4
Neither method is valid	2
Peer assessment measures potential better than citation analysis (it is less retrospective)	2
Both methods are subjective	1
Peer assessment is qualitative, citation metrics are quantitative and the two complement one another	3
Citations are less easily manipulated than peer assessments*	1

Peer review and citation analysis may not be independent because Harter & Hooten (2005, p. 268) showed that citation may influence the outcome of peer review

Table 8. Categorisation of comparisons of peer assessment and citation analysis from qualitative survey data

Peer assessment preferred	Citation analysis preferred	Both methods should be used together	Neither method should be used, an alternative should be found
25	1	19	5
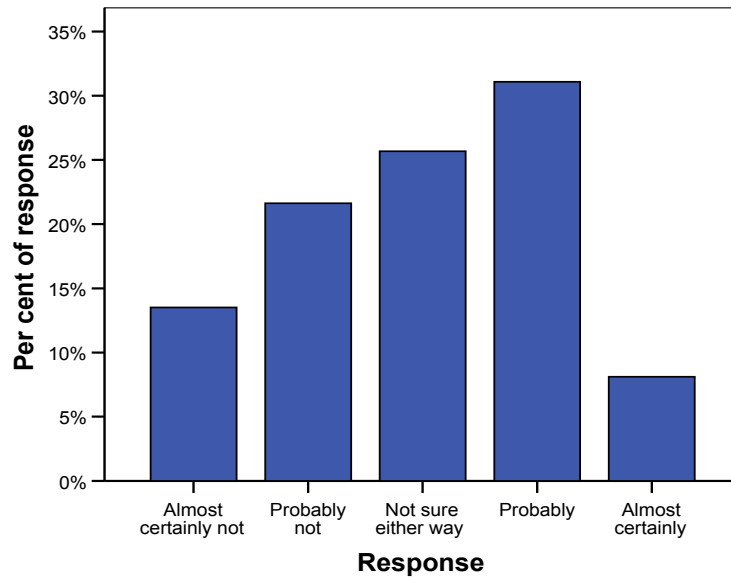
Figure 60. Bar chart showing favourability toward using citation metrics to assess researchers in their own team if citation metrics were shown to correlate reliably with peer assessment

Table 8 suggests that the majority of researchers opposed assessment by citation analysis. Figure 60 further suggests that two-thirds of the survey respondents, who admittedly are not necessarily representative of the underlying population because of the limited response rate, would be at best hesitant to use citation metrics, even if strong evidence were provided that the two methods produced similar outcomes.

4.9 Factor analysis

4.9.1 Validity and reliability

Only seventy-six responses completed all the factor analysis questions. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was 0.501, barely larger than the critical value of 0.5 required for the factor analysis to proceed. No eigenvalues were negative, so the analysis was valid (Kline 1994, p. 40). Bartlett's Test of Sphericity gave $\chi^2_{v=300}$=597.331; P<0.001, confirming that the component matrix was not an identity matrix, and that the answers to questions were significantly correlated. (*How to Perform and Interpret Factor Analysis using SPSS* [n.d.])

The correlation matrix reproduced from the factors was inaccurate in parts, and the communalities along the diagonal that should have been close to unity varied between 0.563 and 0.807, reflecting the large amount of noise in the data and the variance not explained by the extracted factors (Kline 1994, pp. 39-40).

4.9.2 Factor extraction

The orthogonal rotations *Varimax*, *Equimax*, *Quartimax* and the oblique *Promax* rotation (all with Kaiser normalisation) all yielded very similar solutions. Direct *Oblimin* (with Kaiser normalisation) yielded a different solution, but was difficult to interpret and since this rotation yielded independent (orthogonal) factors, the *Varimax* rotation was interpreted.

Nine factors were found with eigenvalues greater than unity. Each factor explained between 5.3% and 9.6% (1 d.p.) of the total variance in the correlation matrix, together explaining 68.2% (1 d.p.) of the total variance.

Most of the correlation coefficients in the matrix were weak (less than 0.3 in magnitude). The amount of noise in the matrix made it difficult to interpret the factors because most questions had loadings of up to 0.5 on each factor and many had broad, weak loadings across several factors of up to 0.3. The factors are described in Table 9.

Table 9. Extracted factors and the significant motivations for citing such publications loading onto them

Factor	Motivations	Loadings	Factor quality
1	Comprehensive coverage Depth of coverage Comprehensively reviews literature	0.835 0.625 0.726	Comprehensiveness
2	Positive results Conclusions are agreed with by myself Others that I respect agree with publications findings Findings are consistent with other research	0.691 0.666 0.704 0.518	Consensus
3	Author regarded as eminent Author known personally	0.723 0.705	Influence/personal knowledge of author
4	Published outside of UK/EU/US NOT published in UK/US	0.783 0.704	Non-EU/US publication
5	Rigorous Important findings Clear and well written	0.857 0.577 0.590	Quality
6	Aware of which areas are funded Cite funded research Ensure own research is in a well-funded field	0.605 0.801 0.710	Subfield funding
7	Specific to my research need Specific to my research topic	0.777 0.824	Relevance
8	Published in EU, outside of UK	0.842	EU publication
9	Individual prefers to be cited by authors whom they cite	0.765	Desire for recognition by peers in subfield

Table 10. Mean scores for survey questions ranked
from 1 (disagree strongly) to 5 (agree strongly)

	Mean	Standard Error of Mean
Relevance 1	5	0
Relevance 2	5	0
Relevance 3	3	0
Relevance 4	3	0
Author 5	3	0
Author 6	2	0
Author 7	3	0
Author 8	2	0
Author 9	3	0
Rigour 10	3	0
Rigour 11	3	0
Rigour 12	4	0
Impact 13	3	0
Impact 14	2	0
Impact 15	4	0
Funding 16	3	0
Funding 17	2	0
Funding 18	2	0
Geography 19	4	0
Geography 20	3	0
Geography 21	2	0
Quality 22	3	0
Quality 23	3	0
Quality 24	3	0
Quality 25	4	0
Journal 26	4	0
Reputation 27	4	0
Funding 28	3	0

4.9.3 Other observations

Whilst most correlations were weak, most answers to Questions 22 and 23 were strongly correlated (r=0.545; P<0.001), whilst the answers to the two questions had a mean of three (after the possible responses were ranked, representing "unsure") with a standard error of zero. Several respondents suggested[2,3] that they had chosen such a middle value when they only sometimes agreed with the statement. Overall, this suggests that works that are agreed upon by experts are

---

[2] McLeod, J. Email to D.E. Bennett, 03.07.07.
[3] Muir, A. Personal communication, 14.08.07.

generally but not always agreed upon by others, and that controversial works are often cited. Surprisingly, responses to Questions 16-18, which load significantly onto factor 6, were not correlated with answers to Question 28, which indirectly asked whether participants thought that others behaved similarly to themselves because researchers at least claimed not to know, shown by the mean response of three (unsure) to Question 28.

Only the *Promax* rotation, which produced the same factors as *Varimax*, found a weak positive association between personal knowledge of author and publication in the UK/US ($r=0.318$) and a weak negative association between publishing outside the UK/US/EU and clarity of exposition ($r=-0.348$). No association was found between other geographical loadings.

Questions 4, 12 19 and 25, 26 and 27 scored means of 4 (agree slightly), suggesting that depth, rigour and clarity of explanation were sought, although not always found, in cited works. It also showed a bias towards UK/US publications, perhaps because of the database coverage bias reported by (Meho 2007, p. 32).

Questions 6, 8, 12, 14, 17, 18, 21 scored 2, suggesting an aversion to positivism: most participants claimed to rarely cite publications that report positive findings, were written by known individuals or prestigious researchers, or which were in well-funded areas. Few individuals cited works outside of the UK, US and EU.

It was surprising that Question 6 had such a low mean, because Question 27 had a mean score of four. This meant that although few researchers admitted to citing prestigious researchers themselves, they believed that others frequently cited prestigious researchers.

5 Discussion

5.1 Demographics

The following analyses are based upon the unproven assumption that the coverage of the citation databases had no significant effect upon the indices calculated. The analysis should therefore be regarded as indicative. To reliably compare these indices, full publication lists would have to be obtained, lack of coverage noted, and the missing citation data obtained manually.

5.2 Comparison of databases

Both databases excluded many, often the same, individuals. In all comparisons, *ScHolar Index* produced significantly larger metrics (Figures 57 and 58), suggesting that it had more extensive publication coverage than *Web of Science*. All *h*-based indices generated by the two citation databases were significantly correlated (Appendix I), suggesting that coverage biases of the two citation databases affected all researchers to a similar extent but the *h*-index scores for individuals varied unpredictably in both databases (Figure 57). Citation based ranking of researchers should rely on a single, comprehensive database that covered all publications of researchers in that field. *ScHolar Index* may have included more titles, more web publications and/or titles that are more heavily cited by Library and Information Science researchers. This significant difference contrasts with Vanclay (2007, p. 1550), who argued that *h* is robust with respect to database coverage biases.

The strong correlations of total publication and total citation counts between the two databases, together with the lack of a convincing correlation for *v* (Figure 59), suggested that the differences between databases mostly comprised selective inclusion of different journal titles.

Total citation and publication counts correlated between the two databases (Figures 55-56), suggesting that the variation in coverage was subtle and involved the inclusion or exclusion of critical publications that affected the other metrics. Total citation counts vary relatively little with errors in the inclusion of publications compared to other metrics, especially *h*, which may be significantly reduced by the exclusion of certain important publications. Citation counts may be more robust, since they are relatively less affected by all known biases, including database inclusion biases, than *h*.

5.3   Correlation analysis – metrics and meaning

5.3.1 Introduction

For the vast majority of metrics, most data points were clustered towards the origin, suggesting that if used for performance assessment in fields with relatively little research activity, it would be difficult to distinguish between them unless improved database coverage spread individuals out more.

Many highly ranked individuals were found either not to be indexed by the citation indices or to have *h*-index scores of zero.

Despite this, the strong correlation between citation counts and publication counts confirmed findings (*e.g.* Hirsch 2005, p. 16572; Cole & Cole 1973, p. 22) that researchers who have a large research impact, such as Nobel Prize winners, were both more heavily cited and more productive than others. This may have been because different subfields of varying activity were compared.

The results support Vinkler (1986, p. 163) and Wallin (2005, p. 262), that citation metrics are only valid as sub-field indicators and are dependent upon subfield size and activity.

5.3.2 Total citation counts

Figure 5 confirms almost exactly the results of van Raan (2006, p. 496), that *h* and total citation counts are positively and curvilinearly associated, suggesting that regardless of how cited an individual was, *h* was constrained by the difficulty to obtain large numbers of citations over a large number of publications.

As a result of this strong correlation, both total citation counts and *h*-index scores showed similarly large degrees of scatter when correlated with total publication counts (Figures 6 and 27). Since mean citation counts largely depend upon total publication count, it was unsurprising that *h*-index scores showed similar scatter against mean citation counts.

5.3.3 *h*-index scores

Few Library and Information Science researchers have achieved *h*-indices greater than ten, only slightly larger than the median of seven for the field found by Oppenheim (2007, p. 300). The results affirm the assertion that an *h*-index of at least five indicates success in Library and Information Science and an *h*-index of at least thirteen indicates an exceptional individual. This clearly confirms that fields cannot be compared without subject normalisation, since Hirsch (2005, p. 16571) suggested that in Physics, successful researchers could expect to enjoy an *h*-index score of at least twenty. The most notable exceptions in Library and Information Science were researchers in the highly active subfield of Chemoinformatics, such as Peter Willett, who was excluded from the analysis because he was so much more productive and so much more heavily cited than the rest of the sample.

The scatter, particularly the few data points in most figures that lie away from the line, suggest that either systematic database inclusion bias or the random factors proposed by Kostoff (1998, p. 33) may occasionally act in concert to produce unusual results for some individuals. It seems that although citation metrics may be broadly reliable at the organisational level, they are not sufficiently consistent

to reliably compare individuals, although the relationship of error to group size has never been formally investigated. Citation metrics might still be safely used as an inexpensive indicator of particularly strong or weak individual performance that should then be subject to independent peer review to explore reasons for unexpectedly high or low citedness within a subfield.

How subfields could be compared where all researchers involved work in the same organisation, or how institutions could be compared (Appendix J) is still problematic, however.

5.3.4 *h* and the other metrics

Based upon the claim by Egghe (2006, p. 131) that *g* would be more sensitive to higher levels of citation, it was expected that the scatter plot of *h* against *g* (Figure 4) would be curvilinear, flattening with increasing *g*, as individuals with more skewed citation distributions increased at a faster rate than *h*. Figure 4 instead shows a linear relationship with little scatter. *h* and *g* would have been expected to increase co-linearly if individuals citation distributions were not skewed, however, the *h*-ranges for the same individuals ranged widely, from zero up to 200, proving that some researchers' publications received disproportionately more citations than others. It appears that Hirsch's *h*-index and Egghe's *g*-index represent, or are affected by, similar underlying citation factors and biases. As a result, all the trends and associations observed for correlations with *h* were observed for *g*, although the strengths of the associations were often slightly greater for *g* than for *h*.

Figures 19 and 27 appear to show total citation counts and total publication counts were curvilinearly related to *h*, reflecting the constraint of the size of *h* in many cases by limited publication count. *g* was not constrained by the number of heavily cited publications (Appendix K), and therefore if Egghe's *g*-index was used as an exclusive measure of research performance, it might encourage the production of poor quality publications by researchers who have published a limited number of highly cited publications in order to maximise their *g*-index scores, in a similar way that assessment by publication count led to a "deluge of

trivial publications" (Hargens & Schuman 1990, p. 205). A high h-index score requires evenly high citation over a number of works, and therefore avoids this danger. This was confirmed by the decreasing rate of increase in h with increasing h^+ (Figure 10), which suggested that the main factor constraining h was the inability to produce a sufficient number of highly cited papers, rather than a lack of citations to publications that have already been heavily cited.

The plot of h against a (Figure 16) showed a similar slope to h against g (Figure 4), but with much greater scatter because a effectively separated out individuals who shared the same h-index score. Given how effectively a separated individuals with similar h-index scores, it was difficult to accept the criticism of Rousseau ([n.d.], p. 4) that a was overly sensitive to variations in citation within Library and Information Science, although this might well be a valid criticism in more active fields. Whether sufficiently many citations separate different individuals with shared h-index scores to reliably overcome the random noise in the data is unknown because the degree of noise has never been quantified.

s correlated very strongly with median overall citation counts, yet failed to correlate with any metrics other than those reflecting overall citedness. v correlated strongly with h, both median citation counts and s (Figures 13-15, Appendix G). This suggests that the number of highly cited publications produced by researchers is proportional to a researcher's publication count.

Since they were strongly correlated with one another but not with h, neither s nor v may be considered measures of skewness. It was hypothesised instead that because they correlated strongly with median overall citation counts, s and v instead measured the length of the long tail of poorly cited works in citation distributions in terms of citations and publications, respectively, a feature which Egghe (2006, p. 131) criticised h for ignoring.

Λ scores correlated strongly with the metrics that measure citation distribution skewness, h-ranges, h^+ and g (r_S=0.456, 0.709 and 0.582, respectively; P<0.001 for each), but not with h, despite the strong correlation of g and h. The extremely shallow slope of the cluster of points in Figure 21 confirmed that Λ

103

cannot easily distinguish between individuals who produce a few highly cited papers and many uncited works (high h-range) and those who publish works that are evenly cited but do not contribute to h (low h-range).

5.3.5 Skewness coefficients

h correlated strongly with h-range, g-range and h^+ (Figures 8-10). This suggested that h^+ is associated with skewness but the high degree of scatter indicated that either different individuals or different subfields had different citation patterns. Some individuals acquired relatively large h^+ scores without a similar increase in h-range, suggesting that some individuals produced several equally highly cited publications.

5.3.6 Conclusions on the *h*-index

h, h-range and h^+ all report different facets of the citation distributions, which together with total publication count and v may be used to demonstrate the impact that a researcher has had upon a field. h-range and h^+ describe the shape of citation distributions accurately enough to distinguish between individuals of different citedness and citation distribution skewness who have similar h-index scores.

This confirmed that h is a heuristic measure of citedness, and suggested that metrics derived from h will further contribute reliable quantitative information about the distribution of citations amongst a collection of publications, although such metrics can only ever be as reliable as the data used to calculate them.

All median cited values were affected by the long tail of uncited publications. Medians were considered unreliable measures because they ignore half the data.

The scatter in all the scatter plots suggests that although the trends were generally consistent, that the associations were too unpredictable to be reliable for the comparison of individuals. These metrics also need to be modified to control for systematic biases, such as the career age of researchers. Without

direct control over this systematic citation bias, citations will remain unreliable measures of individual publication impact.

5.4 Comparison of academic ranks

Figures 32-36 consistently show that both citation metrics such as h and citation distribution skewness metrics such as h^+ and h-range increased with increasing academic rank up to Reader, but then decreased amongst Professors and Heads of Departments. For all academic ranks up to Reader, the results confirmed the findings of Harter & Hooten (1990, p. 268), that citations are important determiners of the outcome of peer assessment, but seemed to suggest that, in the past at least, Professors may have been selected on some other basis than their ability to produce heavily cited publications.

The Readers sampled might have been more heavily cited than the Professors because the Readers included worked in more active subfields, or even different fields, such as Computer Science, or it may have been that both citation databases excluded the main titles in which the Professors had published.

No normalisation was conducted for career age because career ages were not available.

Differences in publication format, *i.e.* the publication of books rather than journal articles may have made a substantial difference to differences in the rankings, especially where it was combined with other factors, such as young career age and restricted range of publication topic[4], a factor probably correlated with age. Finally and most importantly, comparing the citation metrics of individuals across different subfields was possibly unsound without adequate subfield normalisation.

---

[4] R.J. Hartley. Email to D.E. Bennett, 28.06.07.

The inclusion of scientists in fields other than Library and Information Science in the sampling frame makes reliable interpretation of the peer-assessment extremely difficult for researchers in a minority of institutions.

5.5   Comparison of departments

The metrics of researchers in departments awarded different RAE (2001) grades were not significantly different, nor were researchers' metrics correlated with the RAE (2001) of their employing department (Appendix L).

Asknes (2006, p. 174) reported similarly weak associations between citation rankings and RAE grades. The strong correlation found between citation counts and RAE grades found by Norris & Oppenheim (2003) for Archaeology and by Seng & Willett (1995) for Library and Information Science suggested that database coverage bias did not affect the results. The sample examined in this study had a median h-index score of only two, compared to seven for the entire field (Oppenheim 2007, p. 300).

The lack of correlation between RAE grade and citation counts for a sample that under-represented moderately highly cited researchers despite strong correlations being found whenever entire departments were analysed, suggests that the citation counts of these moderately cited individuals determined the outcome of the correlations, and suggests that complete publications lists must be used for assessments.

5.6   Comparisons of peer and *h* based rankings

Peer assessment often disagreed with the h based rankings of individuals, confirming previous research by Asknes (2006, p. 174), and citation metrics were not associated with academic rank, suggesting that promotion and esteem of Library and Information Science researchers in the UK was not related to the citedness of researchers, in conflict with the outcome of the factor analysis by

Harter & Hooten (1990, p. 268). The association of *h* with peer-rankings was no stronger for Professors and Readers than for Lecturers and Research Associates, further suggesting that those who conduct peer assessments assess factors other than previous production of heavily cited research.

It is possible that peer-rankings and academic ranks were associated with the citation metrics studied but the association were obscured by the coverage biases of the citation databases used to calculate the metrics. Library and Information Science is also a small field, so the size of researchers' *h*-index scores were restricted (Garfield & Welljams-Dorof 1992, p. 322), resulting in so many researchers sharing the same *h*-index score. Peer-ascribed ranks were not allowed to be tied, whilst *h*-index ranks often were, thereby unfairly reducing the size of some correlation coefficients. Finally, the response was small compared to the size of the underlying population, and so cannot be reliably generalised.

Many researchers commented that peer assessment often demanded assessors to comment on areas of research they were unfamiliar with (Appendix F), and Asknes (2006, p. 169) proved that self-assessments of publications correlated with citations whilst assessments by others did not. Together with participant feedback[5], this suggested that a lack of expertise in all areas researched by each individual in a department might also have made the peer-rankings less reliable.

5.7   Comparison of genders

5.7.1 Comparison of academic ranks held

Although male researchers tend to be promoted to higher positions slightly more frequently than female researchers, the difference was shown not to be significant, nor was there any association between variations in the female:male rank ratio and the appraisal of institutional research performance.

---

[5] Hartley, R.J. Email to D.E. Bennett, 28.06.07.

Figures 28-31 suggest that male researchers were more heavily cited than female researchers, that more men produced very heavily cited papers, shown by their higher mean h–ranges, and that female researchers produced a greater proportion of poorly cited research, shown by their higher v scores. The standard errors of the means were so large that none of these conclusions can be considered reliable, however.

In contrast to Ferber (1986, p. 382) and Bayer & Austin (1968 *cited in* Cole 1979, p. 57), who found that heavily cited female researchers appeared to have been systematically denied professorships and tenure in 1970s America, current differences observed in the UK appear to reflect the availability of talented male and female researchers when positions of seniority became available.

5.7.2 Peer rankings – gender analysis

The observed and expected frequencies of the two genders (Table 4) were almost identical, confirming that, in addition to the lack of substantial gender bias in the academic rank held by different genders, there was no gender bias in the prestige of highly regarded researchers.

5.8 Qualitative analysis

The distrust of citation metrics was mostly grounded in common perceptions, with criticisms echoing those in the literature. The most frequently raised objections were that negational citations confused the meaning of citation metrics, that citations reflected visibility, and that therefore metrics were retrospective. Both peer review (considered to be part of peer assessment) and citation metrics were felt to exert a "Matthew effect", favouring small groups of well cited researchers and hindering the rise of promising young researchers and improving research groups, but peer review was considered to be more responsive and able to detect and appraise recent change.

Where funding in a field may all be given to one research group metrics that lack currency because of delays in adding citations to databases and delays due to visibility effects may be less useful as means of assessment, even at the departmental level. Concerns over the repression of novel methodologies and ideas and the enforcement of orthodox thinking appeared to be similarly enforced through the two methods.

Figure 59 suggested that many were motivated by fear of change and a distrust of what was perceived as crude quantitative assessment. Several responses to Question 31 also expressed an emotional preference for the more familiar peer assessment, which suggested that it would be difficult to persuade researchers in Library and Information Science to widely adopt citation metrics as a means of assessment, even though RAE assessments have been repeatedly shown to rank departments in the same order as departmental citation counts across different fields (*e.g.* Seng & Willett 1995, p. 70-71; Norris & Oppenheim 2003, p. 727; Oppenheim 1996, p. 160). The comments all closely resembled the misgivings expressed in a heavily cited paper by Warner (2000, p. 455).

The suggestion by one respondent of reviewing publication outcomes qualitatively echoed that of Seglen (1992, p. 636) that "novelty, solidity and magnitude" should be appraised. Certainly novelty is reportedly stifled by both peer assessment and citation analysis. Whether magnitude is always recognised given the lack of knowledge of some reviewers is uncertain. Solidity correlated well with citation counts because unsound works have been shown to attract few, if any, negational citations within (Oppenheim 1996, p. 158) despite several anecdotal accounts of negational citations.

5.9 Factor analysis

The factor analysis confirmed the existence of all proposed factors (Appendix E) and illuminated the relation of these motivations to one another.

The most consistently highly-scored items in the factor analysis (Table 10) represented the Mertonian norms of publication relevance, quality and rigour. The component matrix (Appendix M) shows that these social norms, particularly publication quality and rigour loaded moderately (c.+0.3 loadings) on to most other psychological and political factors. This made it difficult to recognise when simple structure had been achieved.

The moderate loadings of the Mertonian motivations onto all factors together with their high mean values showed that they formed an overarching factor that influenced almost all citation behaviour. That the factors were almost totally independent of one another suggested that the other factors represented motivations that psychological and political motivations, such as desire to agree with a scholarly consensus (factor 2), which Ghaebi (2003, p. 133) mistakenly described as "quality", affect individuals to different extents independently of one another, and that few individuals share the same combination of motivations (including strong aversion to an activity).

Whether a work had been heavily cited previously was claimed not to affect the probability of its being cited (Question 13 had a mean of three). This suggested that although visibility through citation increases the probability of a publication being discovered (*e.g.* Asknes 2003, p. 168; Garfield & Welljams-Dorof 1992, p. 323; Appendix F), visibility was thought not to affect cognitive relevance judgements, nor decisions of whether to cite research, despite the desire of a minority of the respondents to cite publications that contain information other scholars had agreed with.

Influence effects were a factor for a minority of individuals. In some cases, known individuals – which respondents suggested included friends, mentors, students, colleagues and potential reviewers – had been cited for political

110

reasons. Despite these effects, h varied relatively little across the field, suggesting that influence effects were small. If prestigious researchers had attracted a disproportionate number of citations they would have achieved much larger h values than were observed. This confirmed both the assertion of White (2004, p. 110) that most research cites well-known but not exceptional individuals and that whilst potential reviewers may be cited, they are not cited much more often than other well-known researchers. Survey participants maintained their belief that because they cited prestigious researchers rarely and that prestigious researchers are cited more often than most, that others must have cited them heavily.

If most citations are to works in the same subfield, then factor nine suggests that individuals were concerned primarily with their acceptance by other experts (potential reviewers) in their own subfield, indicated by citation frequency. This might also suggest an (unconscious) tendency towards reciprocal citation, which was suggested by several survey participants (Appendix F).

Only just over half of the variance was explained by the factors. This might suggest that other factors might influence citation behaviour that were not tested, or it might have resulted from the tendency of many researchers to give an answer of three, meaning "unsure" (Appendix M).

Whether motivations vary with publication type, researcher, and/or subfield, it appears that they are not consistent across Library and Information Science, and that different researchers have different combinations of major citation motivations. Unless these motivations prove to vary with subfield, citations appear to be a measure of at least nine different, independent factors.

When the factors were initially extracted, it was expected that quality and funding, or quality and relevance would be correlated together to form secondary factors indicating that high quality work was rewarded or that only relevant, high quality work was considered suitable for citation, respectively. Neither was correlated.

In distinct contrast to Brooks (1986), who found that persuasiveness and individual consensus with the published work determined whether it was cited, this study suggests that different researchers in the UK are much more independent of the opinions of others and consider a wider and less predictable range of factors when deciding to cite research. Whether this finding was an artefact of the questions asked is not known, but the lack of correlation between "consensus" and "quality" (which included robustness) suggested that the results of the two studies were substantially different.

Most UK Library and Information Science researchers claimed to cite disproportionately more English language (especially US/UK) research than European and worldwide publications. The high factor loadings and low mean values for non-EU/US/UK publication use suggests a small number of researchers with geographically specific research interests might have existed.

Interestingly, frequent citation of EU (non-UK) research was independent of whether participants often cited research in the UK/US or in other parts of the world, except that those who cite UK/US research tend not to cite research published in other countries, and *vice-versa*. Researchers who cited research outside the UK/US therefore appeared to form a small (West & McIlwaine 2002, p. 503; Pasterkamp *et al.* 2007, p. 163) but distinct group.

It appears to have become a generally accepted truth that most researchers cite peer reviewed articles but altering research focus to ensure continued funding is not. Those who tracked funding knew that they were a minority.

6 Conclusions and recommendations

6.1 Conclusions

6.1.1 Citation analysis and peer assessment

- Both *Web of Science* and *Google Scholar* (via *ScHolar Index*) omit individuals and publications, systematically and arbitrarily biasing citation metrics calculated using them

- It is therefore unlikely that any attempt to reduce random citation biases, such as funding peer review to more closely inspect and prevent miscitations, could render citation metrics accurate measures of research performance

- In contrast to studies in other fields, neither RAE (2001) grades nor other measures of peer assessment, including academic rank, correlated with individuals citation counts or other metrics. Older studies summed citations across whole departments, whilst this study examined individuals. The studies were not entirely comparable, since other studies summed citations across entire departments.

- Citations will always be subfield-specific and retrospective measures of research performance because they rely upon visibility effects. They should therefore be standardised for career age and subfield before comparisons are made between institutions or individuals

- Developing the ideas of Oppenheim (1997, p. 495), the fairest alternative to the combined use of citation analysis with peer assessment appears to be to use existing citation databases to calculate metrics as within-subfield measures to identify those researchers and departments that appear to be cited unusually little or much for their subfield, and subject only these individuals/departments to peer assessment

- Due to the retrospective nature of citation analysis, this method would (dis)advantage institutions whose staff or performance had recently changed

- Citation analysis alone would fail to provide the meaningful, rich feedback that follows peer assessment

- Resistance to purely citation-based assessment is widespread and fierce

- Citation metrics are distrusted because of their imprecise meaning and because of concern that the retrospective nature of citations might delay the promotion of researchers of younger career age

6.1.2 Comparison of citation metrics

- The large quantity of noise in the citation data gathered might reflect database publication bias or different citation patterns in different subfields

- *h* must be a heuristic measure because it was possible to extract so many metrics that measure different citation distribution properties from it. Together, these metrics may give a reliable indication of the citation distribution of an individual or department

- *g* was only slightly more sensitive than *h* for the data set analysed, despite many researchers having highly skewed citation distributions, perhaps because the *g*-index scores of many individuals were constrained by low publication counts

- Both the *h* and *g* scores of many others were constrained by total publication count

- Median citation counts ignore too much of the data to be considered reliable

- *a* effectively distinguished between researchers with identical *h*-index scores, therefore if the number of citations that separates individuals with different *a* values is larger than that likely to be accrued through miscitations, *a*-based metrics would be more useful than *h*-based metrics in small fields with relatively low levels of research activity

- *s* and *v* measure the number of publications which are poorly cited relative to the size of *h*. They increase with the number of citations and publications,

respectively, that do not contributed to *h*. They would be equally meaningful if *a* was substituted for *h*

- *h* may be of less use than total citation counts in Library and Information Science because *h* is constrained by the low activity of the field and therefore many researchers share the same *h*-index score, whereas productivity and citedness varied more widely, allowing individuals to be separated out although how large a difference would be needed before a gap could be regarded as reliable in this field is unknown

- Productivity was only weakly associated with citation counts, suggesting that either researchers of different productivities produce work of equally great interest to other researchers or that equally productive researchers working in different subfields were cited to different extents

- Many scatter plots showed deviating trends that might have represented subfields that did not obey general trends or were misrepresented because of database coverage bias

6.1.3 Academic ranks and citation metrics

- The order in which individuals are ranked by their *h*-index scores varies depending upon which citation database is used to calculate the metric. Therefore metric scores and rankings calculated using data from different databases cannot be compared.

- Academics of higher rank published more very highly cited publications, represented by the increase of median *h*-ranges with increasing rank

- Professors, whether because of database coverage bias or because they were promoted on some other basis than the ability to publish highly cited journal articles, had much lower metrics than some researchers of lower academic rank

6.1.4 Peer ranking and citation metrics

- Peer-rankings appeared to measure something different from productivity or citedness because neither academic rank nor peer-rankings correlated with *h*-based rankings

- No gender bias was evident in academic rank, esteem or citedness

6.1.5 Factor analysis

- Nine factors explained over half of the variance in declared citation motivation: comprehensiveness, consensus, author influence, geographical location, quality, subfield funding, relevance, and desire for peer recognition

- Mertonian scientific norms were claimed to influence all citation behaviour, suggesting that, given the reported scarcity of negational citations, citation metrics measure both quality and visibility/impact

- The independence of the factors suggested that different researchers had different (combinations of) political and psychological motivations for citing research, and that few researchers shared similar combinations of motivations

- Few researchers admitted to tracking research funding with their research focus. Those who did were aware that they were a minority

- Most researchers cited prestigious researchers infrequently but mistakenly believed that other researchers must cite them frequently in order for them to be heavily cited

6.2 Recommendations for future research

- Quantify the effects of suspected biases and whether they act systematically or randomly (Appendix O)

- Assess whether random biases do act randomly, the amount that they are likely to alter total citation counts and the probability that they will alter h, for any individual

- Develop a robust method to identify subfields (Appendix P)

- Develop a simple, robust method for subfield citation normalisation. Subfields might then be combined such that the metric scores of each individual in the subfield might be subtracted from the subfield median, the differences normalised to make subfields comparable, and summed to produce comparable statistics for different research groups, departments and institutions (Appendix J)

- Investigate whether the degree to which citation metrics are retrospective varies between subfields

- Investigate how citation metrics change over time, and whether they significantly increase or decrease for individual researchers or across fields and subfields over time, showing halo effects (Cronin & Crawford 1999, p. 473) and whether subfield growth rates are correlated with rates of publication obsolescence, as observed for productivity and article citation half-lives in journals (Nicholas *et al*. 2005, p. 1443-1444)

- Compare citation metrics with the total number of times a work is cited in subsequent works, to identify whether the latter measures, proposed by White (2004, p. 87), are distorted by citation styles within subfields

- Compare citations and acknowledgements with citations alone as measures of esteem and reward

- Study the rate at which citations are accumulated by different types of publication by authors of different prestige and productivity across different

subfields with different research activity rates, to see whether publication visibility increases linearly or whether a threshold number of citations exists for different publications after which a work becomes highly visible and whether the effect varies between subfields

- Metrics for different academic ranks should be re-examined within, and compared between, subfields

- Citation metrics may be used in other areas of Library and Information Science, such as borrowing in collection management, and in other fields where a particular action by population members signifies relative approval (Appendix N)

- Repeat the correlation of RAE (2001) and, when available RAE (2008) with the publication and citation data for all individuals in all departments as submitted for the assessments

Bibliography

Ahmed, T., *et al.*, 2004. Highly cited old papers and the reasons why they continue to be cited. Part II: the 1953 Watson and Crick article on the structure of DNA. *Scientometrics* [online], 61 (2), 147-156. <http://www.springerlink.com/content/m64786up815242r1/fulltext.pdf>, [accessed 28.08.07].

Asknes, D.W., 2003. Characteristics of highly cited papers. *Research Evaluation*, 12 (3), 159-170.

Asknes, D.W., 2006. Citation rates and perceptions of scientific contribution. *Journal of the American Society for Information Science and Technology* [online], 57 (2), 169-185. <http://www3.interscience.wiley.com/cgi-bin/fulltext/112141120/PDFSTART>, [accessed 29.08.07].

Ashton, S.V. & Oppenheim, C., 1978. A method of predicting Nobel prizewinners in chemistry. *Social Studies in Science*, 8 (3), 341-348.

Baird, L.M. & Oppenheim, C., 1994. Do citations matter? *Journal of Information Science* [online], 20 (2), 1-15. <http://jis.sagepub.com/cgi/reprint/20/1/2>, [accessed 28.08.07.].

Bar-Ilan, J., 2006. An ego-centric citation analysis of the works of Michael O. Rabin based on multiple citation indexes. *Information Processing & Management* [online], 42 (6), 1553-1566. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC8-4K07N5C-1&_user=122878&_coverDate=12%2F31%2F2006&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=d7439424dffa1cd2c2bceec20a48f7d0>, [accessed 30.08.07].

Batista, P.D., *et al.*, 2005. An index to quantify an individuals scientific research valid across disciplines. *Scientometrics* [online], 68 (1), 179-189. <http://www.springerlink.com/content/a41351w08632151v/fulltext.pdf>, [accessed 28.08.07].

Bornmann, L. & Daniel, H.-D., 2007. What do we know about the *h*-index. *Journal of the American Society for Information Science and Technology*, 58 (9), 1381-1385.

Bornmann, L. & Daniel, H.-D., 2006. Selecting scientific excellence through committee peer review – a citation analysis of publications previously published to approval or rejecting of post-doctoral research fellowship applicants. *Scientometrics*, 18 (3), 427-440.

Bornmann, L. & Daniel, H.-D., 2005. Does the *h*-index for ranking scientists really work? *Scientometrics*, 65 (3), 391-392.

Bourque, L.B. & Fielder, E.P., 2003. *How to conduct self-administered and mail surveys*, 2nd ed. Thousand Oaks; London: Sage.58(9)

Bowker, D.W. & Randerson, P.F., 2005. *Minitab workbook for biosciences*, Minitab 14 ed. Cardiff: Cardiff School of Biosciences.

Boyack, K.W. & Börner, K., 2003. Indicator-assisted evaluation and funding of research. *Journal of the American Society for Information Science and Technology* [online], 54 (5), 447-461. <http://www.springerlink.com/content/v446nt7553280755/fulltext.pdf>, [accessed 20.04.07].

Braun, T., Glänzel, W. & Schubert, A., 2000. How balanced is the *Science Citation* Index's journal coverage? A preliminary overview of macrolevel statistical data. Cronin, B. & Atkins, H.B., eds. *The Web of Knowledge: a festschift in honor of Eugene Garfield*, pp. 251-281. Medford: Information Today.

Brooks, T.A., 1986. Evidence of complex citer motiations. *Journal of the American Society for Information Science* [online], 37 (1), 34-36. <http://www3.interscience.wiley.com/cgi-bin/fulltext/114269211/PDFSTART>, [accessed 28.08.07].

Busha, C.H., & Harter, S.P., 1980. *Research methods in librarianship: techniques and interpretation.* New York; London: Academic Press.

Case, D.O. & Higgins, G.M., 2000. how can we investigate citation behaviour? A study of reason for citing literature on communication. *Journal of the American Society for Information Science* [online], 51 (7), 635-645. <http://www3.interscience.wiley.com/cgi-bin/fulltext/72502630/PDFSTART>, [accessed 28.08.07].

Cole, J.R., 1979. *Fair science: women in the scientific community.* London: Free Press.

Cole, J.R. & Cole, S., 1973. *Social stratification in science.* Chicago & London: University of Chicago.

Cole, J.R. & Cole, S., 1977. Letter to *Science. Science*, 183 (4120), 32-33.

Cozzens, S.E., 1988. What do citations count? The rhetoric-first model. *Scientometrics*, 15 (5-6), 437-477.

Cronin, B. & Atkins, H.B., eds., 2000. *The Web of Knowledge: a festschrift in honor of Eugene Garfield.* Medford: Information Today.

Cronin, B. & Crawford, H., 1999. Do deans practice what they preach? *Journal of the American Society for Information Science* [online], 50 (5), 471-474. <http://proquest.umi.com/pqdlink?Ver=1&Exp=04-19-2012&FMT=7&DID=66376084&RQT=309&cfc=1>, [accessed 28.08.07].

Cronin, B., & Meho, L., 2006. Using the *h*-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology* [online], 57 (9), 1275-1278. <http://www3.interscience.wiley.com/cgi-bin/fulltext/112604479/PDFSTART>, [accessed 28.08.07].

Cronin, B., 1984. *The citation process*. London: Taylor Graham.

Cronin, B., 2000. Semiotics and evaluative bibliometrics. *Journal of Documentation* [online], 56 (4), 440-453. <http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Published/EmeraldAbstractOnlyArticle/Pdf/2780560405.pdf>, [accessed 29.08.07].

Cronin, B., 2005. *The hand of science: academic writing and its rewards*. Oxford: Scarecrow.

Daniels, D.E., *et al.*, 2002. Effect of publication bias on retrieval bias. *Academic Medicine*, 77 (3), 266.

de Moya-Anegón, F., *et al.*, 2007. Coverage analysis of *Scopus*: a journal metric approach. *Scientometrics* [online], 73 (1), 53-78. <http://www.springerlink.com/content/q545p23111570447/fulltext.pdf>, [accessed 31.08.07].

Döring, T.F., 2007. Quality evaluation needs some better quality tools. *Nature*, 445 (7129), 709.

Egghe, L., 2006. Theory and practise of the *g*-index. *Scientometrics* [online], 69 (1), 131-152. <http://www.springerlink.com/content/4119257t25h0852w/>, [accessed 28.08.07].

Ferber, M.A., 1986. Citations: are they an objective measure of scholarly merit. *Signs: Journal of Women in Culture and Society*, 11 (2), 381-389.

Fink, A., 1995. *How to ask survey questions*. Thousand Oaks; London: Sage.

Fowler, F.J., 2002. *Survey research methods*, 3rd ed. Thousand Oaks; London: Sage.

García-Aracil, A., Gracia, A.G. & Pérez-Marín, M., 2006. Analysis of the evaluation process of the research performance: and empirical case. *Scientometrics* [online], 67 (2), 213-230. <http://www.springerlink.com/content/lxp3584517jx1671/fulltext.pdf>, [accessed 28.08.07].

Garfield, E. & Welljams-Dorof, A., 1992. Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. *Science and Public Policy*, 19 (5), 321-327.

Garfield, E., 1997. Validation of citation analysis. *Journal of the American Society for Information Science* [online], 48 (10), 962-964. <http://www3.interscience.wiley.com/cgi-bin/fulltext/39610/PDFSTART>, [accessed 29.08.07].

Garfield, E., 1998. From citation indexes to informetrics: is the tail now wagging the dog? *Libri*, 48 (2), 67-80.

Getz, M., 2005. Open-access scholarly publishing in economic perspective. *Journal of Library Administration* [online], 42 (1), 1-39. <http://www.haworthpress.com/store/E-Text/View_EText.asp?sid=091TGACMRMC98MVQXLB0DKA9KVGEF1C7&a=3&s=J111&v=42&i=1&fn=J111v42n01%5F01>, [accessed 22.05.07].

Ghaebi, A., 2003. *Measurement of factors determining relevance judgements*. PhD thesis, Loughborough University.

Glänzel, W. & Schubert, A., 2001. Double effort = double impact? A critical view at international co-authorship in chemistry. *Scientometrics* [online], 50 (2),

199-214. <http://www.springerlink.com/content/t2pq1372401v27v8/>, [accessed 28.08.07].

Gorman, G.E. & Clayton, P., 2005. *Qualitative research or the information professional: a practical handbook*, 2nd ed. London: Facet.

Great Britain: HM Treasury, 2006. *Science and innovation investment framework 2004-2014: next steps*. <http://www.hm-treasury.gov.uk/media/7/8/bud06_science_332v1.pdf>, [accessed 28.08.07].

Hargens, L.L. & Schuman, H., 1990. Citation counts and social comparisons: scientists use and evaluation of citation index data. *Social Science Research* [online], 19 (3), 205-221. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6WX8-4D5W378-J&_user=122878&_coverDate=09%2F30%2F1990&_alid=613706125&_rdoc=27&_fmt=summary&_orig=search&_cdi=7152&_sort=d&_docanchor=&view=c&_ct=160&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=457c47c4479e029046670c6f0ff2616c>, [accessed 13.06.07]

Harnad, S., *et al*., 2004. The access/impact problem and the green and gold roads to open access. *Serials Review* [online], 30 (4), 310-314. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W63-4DS8FN9-F&_user=122878&_coverDate=12%2F31%2F2004&_alid=613650467&_rdoc=3&_fmt=full&_orig=search&_cdi=6587&_sort=d&_docanchor=&view=c&_ct=5&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=cf477f2bdc4a9994f2de9715ff082858>, [accessed 29.08.07].

Hart, R. L., 2007. Collaboration and article quality in the literature of academic librarianship. *The Journal of Academic Librarianship*, 33 (2), 190-195.

Harter, S.P. & Hooten, P.A., 1990. Factors affecting funding and citation rates in information science publications. *Library and Information Science Research*, 12 (3), 263-280.

Harter, S.P., 1992. Psychological relevance and information science. *Journal of the American Society of Information Science* [online], 43 (9), 602-615. <http://www3.interscience.wiley.com/cgi-bin/fulltext/10049714/PDFSTART>, [accessed 21.04.07].

Heath, H. & Cowley, S., 2004. Developing a grounded theory approach: a comparison of Glaser and Strauss. *International Journal of Nursing Studies* [online], 41, 141-150. <http://www.sciencedirect.com/science?_ob=PublicationURL&_tockey=%23TOC%235067%232004%23999589997%23475959%23FLA%23&_cdi=5067&_pubType=J&view=c&_auth=y&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=4bb649b0c1f035a632ca3cec9b77c293>, [accessed 11.04.07].

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences* [online]*, 102 (46), 16569-16572. <http://www.pnas.org/cgi/reprint/102/46/16569>, [accessed 28.08.07].

Holden, G., 2006. An assessment of the predictive validity of impact factor scores: implications for academic employment decisions in social work. *Research on Social Work Practice*, 16 (6), 613-624.

How to perform and interpret factor analysis using SPSS, [n.d.]. <http://www.ncl.ac.uk/iss/statistics/docs/factoranalysis.html>, [accessed 28.07.07].

Janes, J., 1999. Survey construction. *Library Hi-Tech* [online], 17 (3), 321-325. <http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2380170312.pdf>, [accessed 11.04.07].

Jarneving, B., 2005. A comparison of two bibliometric methods for mapping of the research front. *Scientometrics* [online], 65 (2), 245-263. <http://www.springerlink.com/content/g2637g81m5721245/fulltext.pdf>, [accessed 30.08.08].

Kelly, C.D. & Jennions, M.D., 2006. The h index and career assessment by numbers. *Trends in Ecology and Evolution* [online], 21 (4), 167-170. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VJ1-4J6241H-1&_user=122878&_coverDate=04%2F30%2F2006&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=fd7b0f9b61f20807aa417cae23a9e188>, [accessed 28.08.07].

Kelt, M., 1996. Some quick tips on survey work. *Assistant Librarian*, 89 (10), 161-163.

Kim, K., 2004. The motivation for citing specific references by social scientists in Korea: The phenomenon of co-existing references. *Scientometrics* [online], 59 (1), 79-93. <http://www.springerlink.com/content/w72867763h4058rm/fulltext.pdf>, [accessed 29.08.07].

Kinnear, P.R. & Gray, C.D., 2006. *SPSS 14 made simple*. Hove: Psychology Press.

Kline, P., 1994. *An easy guide to factor analysis*. London: Routledge.

Kochran, C.A. & Budd, J.M., 1992. The persistence of fraud in the literature: the Darsee case. *Journal of the American Society for Information Science* [online], 43 (7), 488-493. <http://www3.interscience.wiley.com/cgi-bin/fulltext/10049692/PDFSTART>, [accessed 29.08.07].

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

Kostoff, R.N., 2002. Citation analysis of research performer quality. *Scientometrics*, 53 (1), 49-71.

Leydesdorff, L. & Amsterdamska, O., 1990. Dimensions of citation analysis. *Science, Technology & Human Values*, 15 (3), 305-335.

Liang, L., 2006. *h*-index sequence and *h*-index matrix: constructions and applications. *Scientometrics* [online], 69 (1), 153-159. <http://www.springerlink.com/content/k221240t574l1t7t/fulltext.pdf>, [accessed 29.08.07].

Liu, M., 1993*a*. A study of citing motivation of Chinese scientists. *Journal of Information science*, 19 (1), 13-23.

Liu, M., 1993*b*. The complexities of citation practice: a review of citation studies. *Journal of Documentation*, 49 (4), 370-408.

MacRoberts, M.H. & MacRoberts, B.R., 1986. Quantitative measures of communication in science: a study of the formal level. *Social Studies of Science*, 16 (1), 151-172.

MacRoberts, M.H. & MacRoberts, B.R., 1987*a*. Another Test of the Normative Theory of Citing. *Journal of the American Society for Information Science* [online], 38 (4), 305-306. <http://www3.interscience.wiley.com/cgi-bin/fulltext/114269408/PDFSTART>, [accessed 29/06.07].

MacRoberts, M.H. & MacRoberts, B.R., 1987*b*. Citation counts. *Nature*, 327 (6122), 456.

MacRoberts, M.H. & MacRoberts, B.R., 1996. Problems of citation analysis. *Scientometrics* [online], 36 (3), 435-444.

<http://www.springerlink.com/content/q324544162u20u56/fulltext.pdf>, [accessed 28.07.08].

Manafy, M., 2007. Scopus harnesses the *h*-index to increase the quality and reliability of citation tracking. *Econtent* [online], 30 (3), 10-11. <http://proquest.umi.com/pqdweb?index=0&did=1252796991&SrchMode=1&sid=1&Fmt=6&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1188271188&clientId=5238>, [accessed 28.08.07].

Manly, B.F.J., 1994. *Multivariate statistical analysis: a primer*, 2nd ed. Boca Raton; London: Chapman & Hall/CRC.

Meho, L.I. & Sonnenwald, D.H., 2000. Citation ranking versus peer evaluation of senior faculty research performance: a case study of Kurdish scholarship. *Journal of the American Society for Information Science*, 51 (2), 123-138.

Meho, L.I., 2007. The rise and rise of citation analysis. *Physics World*, 20 (1), 32-36.

Moed, H.F., 1989. *The use of bibliometric indicators for the assessment of research performance in the natural and life sciences*. Leiden: DSWO.

Murray, P., 1999. Fundamental issues in questionnaire design. *Accident and Emergency Nursing* [online], 7 (3), 148-153. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6W9C-4KCRCHM-4&_user=122878&_coverDate=07%2F31%2F1999&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=c81a536e66782bd391bfd382c00188f2>, [accessed 30.08.07].

Nederhof, A.J. & van Raan, A.J.F., 1987*a*. Citation theory and the Ortega Hypothesis. *Scientometrics*, 12 (5-6), 325-328.

Nederhof, A.J. & van Raan, A.J.F., 1987*b*. Peer review and bibliometric indicators of scientific performance: a comparison of *cum laude* doctorates with ordinary doctorates in physics. *Scientometrics*, 11 (5-6), 333-350.

Nicholas, D., *et al.,* 2005. Revisiting obsolescence and journal article decay through usage data: an analysis of digital journal use by year of publication. *Information Processing and Management* [online], 41 (6), 1441-1461. <http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC8-4GPW3S3-1&_user=122878&_coverDate=12%2F31%2F2005&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=2f9731836fdbcf7148d5ad6e3d709bae>, [accessed 28.08.07].

Norris, M. & Oppenheim, C., 2003. Citation counts and the research assessment exercise V: archaeology and the RAE. *Journal of Documentation* [online], 59 (6), 709-730. <http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2780590605.pdf>, [accessed 31.08.07].

Oppenheim, A.N., 1992. *Questionnaire design, interviewing, and attitude measurement*, 2nd ed. London: Printer.

Oppenheim, C., 1996. Do citations count? *Serials* [online], 9 (2), 155-161. <http://www.metapress.com/content/5ycdb0m2k3xgaya6/fulltext.pdf>, [accessed 22.08.07].

Oppenheim, C., 1997. The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British research in Genetics, Anatomy and Archaeology. *Journal of Documentation* [online], 53 (5), 477-487. <http://www.emeraldinsight.com/Insight/viewPDF.jsp?Filename=html/Output/Published/EmeraldAbstractOnlyArticle/Pdf/2780530502.p>, [accessed 15.08.07].

Oppenheim, C., 2006. *Case for support: using the h index to rank influential UK researchers in information science and librarianship.* Unpublished.

Oppenheim, C., 2007. Using the *h*-index to rank influential British researchers in information science and librarianship. *Journal of the American Society for Information Science and Technology* [online], 58 (2), 297-301. <http://www3.interscience.wiley.com/cgi-bin/fulltext/113489425/PDFSTART>, [accessed 29.08.07].

Pasterkamp, G., *et al*., 2007. Citation frequency: a biased measure of research impact significantly influenced by the geographical origin of research articles. Scientometrics [online], 70 (1), 153-165. <http://www.springerlink.com/content/vh78876tru122k43/fulltext.pdf>, [accessed 28.08.07].

Po, A.L.W., 1997. *Statistics for pharmacists*. Oxford: Blackwell.

Powell, R.R., 1985. *Basic research methods for librarians*. Norwood, NJ: Ablex.

Rousseau, R., [n.d.]. New developments related to the Hirsch index. [Available from *ArXiv*]. <http://eprints.rclis.org/archive/00006376/01/Hirsch_new_developments.pdf >, [accessed 14.06.07].

Saad, G., 2006. Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics* [online], 69 (1), 11-120. <http://www.springerlink.com/content/p5077ln623g3/>, [accessed 28.08.07].

Schuman, H. & Presser, S., 1981. *Questions and answers in attitude surveys: experiments on question form, wording, and context.* New York; London: Academic Press.

Schwartz, C.A., 1997. The rise and fall of uncitedness. *College and Research Libraries* [online], 58 (1),19-29.

<http://www.ala.org/ala/acrl/acrlpubs/crljournal/backissues1997b/january97/schwartz.pdf>, [accessed 29.08.07].

Seglen, P.O., 1992. The skewness of Science. *Journal of the American Society for Information Science* [online], 43 (9), 628-638. <http://www3.interscience.wiley.com/cgi-bin/fulltext/10049716/PDFSTART>, [accessed 28.08.07].

Seng, L.B. & Willett, P., 1995. The citedness of publications by United Kingdom library schools. *Journal of Information Science*, 21 (1), 68-71.

Siegel, S. & Castellan, N.J., 1988. *Nonparametric statistics for the behavioural sciences*, 2nd ed. London: McGraw-Hill.

Simkin, M.V. & Roychowdhury, V.P., [n.d.]. Read before you cite! [Available from *ArXiv*]. <http://arxiv.org/ftp/cond-mat/papers/0212/0212043.pdf>, [accessed 28.07.07].

Sudman, S., Bradburn, N.M., 1982. *Asking Questions.* San Francisco; London: Jossey-Bass.

Sudman, S., Bradburn, N.M., Schwartz, N., 1996. *Thinking about answers: the application of cognitive processes to survey methodology.* San Francisco: Jossey-Bass.

Swigger, K. & Wilkes, A., 1991. The use of citation data to evaluate serials subscriptions in an academic library. *Serials Review* [online], 14 (2), 41-52. <http://www.sciencedirect.com/science?_ob=ArticleListURL&_method=list&_ArticleListID=613857979&_sort=d&view=c&_acct=C000010119&_version=1&_urlVersion=0&_userid=122878&md5=123500a7aebe4444a44e0f28732524f3>, [accessed 29.08.07].

Torgerson, C.J., 2006. Publication bias: the Achilles' heel of systematic reviews? *British Journal of Educational Studies* [online], 54 (1), 89-102. <http://www.blackwell-synergy.com/toc/bjes/54/1>, [accessed 29.08.07].

Trimble, V., 2005. Productivity and impact analysis: rediscovering the obvious. *Scientometrics* [online], 63 (2), 413-416. <http://www.springerlink.com/content/g272213115ng2112/fulltext.pdf>, [accessed 28.08.07].

van Leeuwen, T.N., Moed, H.F., Reedjik, J., 1999. Critical comments on Institute for Scientific Information impact factors: a sample or inorganic molecular chemistry journals. *Journal of Information Science* [online], 25 (6), 489-498. <http://jis.sagepub.com/cgi/reprint/25/6/489>, [accessed 28.08.07].

van Raan, A.F.J., 2006. Reference-based publication networks with episodic memories. *Scientometrics*, 63 (3), 549-566.

van Raan, A.F.J., 2005*a*. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62 (1), 133-143.

van Raan, A.F.J., 2005*b*. Reply to the comments of Liu *et al*. *Scientometrics*, 64 (1), 111-112.

van Raan, A.F.J., 2000. The Pandora's Box of citation analysis: measuring scientific excellence the last evil? *In:* Cronin, B. & Atkins, H.B., eds. *The Web of Knowledge: a festschrift in honor of Eugene Garfield*m pp. 301-320. Medford: ASIS.

Vanclay, J.K., 2007. On the robustness of the *h*-index. *Journal of the American Society for Information Science and Technology*, 58 (10), 1547-1550.

Vinkler, P., 1986. Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics*, 10 (3-4), 157-177.

Vinkler, P., 1987. Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics*, 10 (3-4), 157-177.

Vinkler, P., 2007. Eminence of scientists, quality against quantity of information. *Journal of Information Science,* in press.

Wade, N., 1975. Citation analysis: a new tool for science adminstrators. *Science*, 188, 429-432.

Wallace, D.P., 1987. A solution in search of a problem: bibliometrics and libraries. *Library Journal*, 112 (8), 43-47.

Wallin, J.A., 2005. Bibliometric methods: pitfalls and possibilities. *Basic and Clinical Pharmacology and Toxicology*, 97, 261-275.

Warner, J., 2000. A critical review of the application of citation studies to the Research Assessment Exercises. *Journal of Information Science* [online], 26 (6), 453-460. <http://jis.sagepub.com/cgi/reprint/26/6/453>, [accessed 15.08.07].

Weingand, D.E., 1993. Grounded theory and qualitative methodology. *IFLA Journal*, 19 (1), 17-26.

West, R. & McIlwaine, A., 2002. What do citation counts in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality. *Addiction* [online], 97, 510-504. <http://www.blackwell-synergy.com/toc/add/97/5>, [accessed 28.08.07].

White, H.D., 2004. Reward, persuasion, and the Sokal Hoax: a study in citation identities. *Scientometrics* [online], 60 (1), 93-120. <http://www.springerlink.com/content/t598765n8655x766/BodyRef/PDF/11192 _2004_Article_5149151.pdf>, [accessed 14.04.07].

White, M.D. & Wang, P., 1997. A qualitative study of citing behavior: contributions, criteria, and metalevel documentation concerns. *Library Quarterly*, 67 (2), 122-154.

Wright, M. & Armstrong, J.S., 18 July 2007. Verification of citations: Fawlty Towers of knowledge? [Available from: *Munich Personal RePEc Archive –* paper N° 4149]. <http:// mpra.ub.uni-muenchen.de/ 4149/>, [accessed 21.04.07].

Yang, K. & Meho, L.I., 23 December 2006. Citation analysis: a comparison of Google Scholar, Scopus, and Web of Science. [Available from *ArXiv*]. <http://arxiv.org/abs/cs.DL/0612132>, [accessed 15.06.07].

Zuckerman, H., 1987. Citation analysis and the complex problem of intellectual influence. *Scientometrics*, 12 (5-6), 329-338.

Appendix A

a

The *a*-index combined the best features of the older *q* index mentioned by Hirsch (2005, p. 16569) and the *h*-index. *a* takes a single value for each researcher because rather than listing the *q* most cited papers, it averages them. By averaging the citations over the *h* most cited papers, *a* is particularly useful for comparing researchers with tied *h*-index scores, regardless of the size of *h*. *h* and *g* only reflect lifetime performances because they take time to increase (Hirsch 2005, p. 16571; Liang 2006, p. 153; Egghe 2006, p. 142). Regardless of the different patterns in research output over researchers lifetimes, because *a* increases with every citation gained, it might be used to compare the impact of research after the first time a researcher was cited. *a* is very difficult to manipulate because it takes into account only the impact of the researchers' most cited publications, and assesses a no more arbitrary a number of publications than *h* does.

h⁺

$$h^+ = \sum_1^h c_i - h^2$$

> = the number of citations contributing to *h* above the minimum number
> required to achieve that value of *h*

h^+ is necessary to differentiate between researchers whose *h* most cited publications are very heavily and evenly cited but whose other publications are cited insufficiently often to achieve an *h*-index score of *h* + 1, and other researchers whose *h*-index score may be higher and whose Λ-index score may also be equal but whose most cited research is much less cited and whose citation distribution is the minimum to achieve that *h*-index score.

g^+

Following the above logic,

$$g^+ = \sum_{1}^{g} c_i - g^2$$

> = the number of citations contributing to g above the minimum number
> required to achieve that value of g

Λ

Λ (after Egghe 2006, p. 143) is proposed as a measure of the number of times the *h* most cited publications have collectedly been cited more than the minimum theoretically necessary to achieve the value of *h* a researcher has attained. Λ is thought to be a measure of the skewness of the distribution of number of citations per publication amongst the *h* most cited publications. If a few citations are cited much more heavily than the rest, *g>h*. If they are all cited the same number of times, *g=h*.

Theoretically, it should be possible to devise a ratio test to test whether differences between *g* and *h* are statistically significant at the 5% significance level, perhaps using the chi-squared distribution.

h_c-range

The range between the most cited paper and the least cited paper contributing to *h*, this is a crude measure of skewness but is calculated from the same data required to discover *h*.

g_c-range

The range between the most cited paper and the least cited paper contributing to *g* this is a crude measure of skewness but is calculated from the same data required to discover *g*.

v

Measures the proportion of a researcher's total publication output that contributes to *h*. See background. It is reasoned that as a researcher continues to publish at a fast rate, and as their *h* index value increases, that unless their work is being cited at an increasing rate, that the proportion of their (recently published) work that does not contributes to *h* would increase, at least until their rate of publication slowed towards the end of their career. Such a trend in productivity was predicted by Hirsch (2005) and observed by Liang (2006).

Reliability of the Λ, *h*-range and *g*-range and median metrics

In less actively researched fields, where papers generally attract fewer citations, "random" citations will have a greater impact, especially since *h* is more vulnerable to 'random' citations – affecting younger researchers and practitioners whose research publication output is least.

Random citation bias that altered *h* would also significantly reduce Λ because

$$\Lambda = \frac{g}{h} \ .$$

However, the *h*-range and *g*-range statistics would only be affected in cases where the paper previously contributing least to *h* or *g*, respectively, had been cited considerably more than the next most cited paper. Since the *h*-range and *g*-range are all supposedly measuring the same thing as Λ, correlations between Λ and these other two statistics should yield $r_S = +1$ if they measure the same thing, and $r_S < 1$ if random effects alter the reliability of Λ. Since a researcher's collection of most highly cited papers contribute more to the value of *g* than does the least cited paper contributing to *h*, the *g*-range should be affected least of all. It is suggested that this is therefore the most reliable index of skewness.

Egghe's *g*-index gives greater credit to highly cited works, which the literature suggests are the pivotal publications that contribute most to the progression of lines of enquiry in science. The *g*-range would be the most reliable measure of skewness because it is robust with respect to (*i.e* it is least affected by) small numbers of perfunctory or spurious citations.

Medians may be even more severely affected by incorrectly recorded, spurious or perfunctory citations to otherwise uncited publications because where these add even single citations to otherwise uncited works, they reduce the median of all cited works (M_c) and increase the value of the overall median (M_o).

Appendix B

Sampling

For any of the results of the survey to be generalisable to the population of UK library and information science researchers, samples must be drawn from the entire population of UK library and information science researchers using a probabilistic sampling method because random, independent sampling is a fundamental assumption of all statistical methods (Siegel & Castellan 1988, p. 34).

Correlation analysis

Most of the citation metrics produce discontinuous data. The most suitable test for association between variables will therefore be Spearman's Rank Correlation Coefficient (SRCC). SRCC is a non-parametric measure of the strength of association between two independent, ranked variables (Siegel & Castellan 1988, p. 235). It is calculated using the formula:

$$r_S = 1 - \frac{6\,\Sigma d^2}{n\,(n^2 - 1)}$$

where r_S = Spearman's Rank Correlation Coefficient

d = difference between ranks for each data point

n = number of data points in correlation

(adapted from Siegel & Castellan 1988, p. 237)

If the ranks for the two independent parameters for each individual data points are similar, SRCC returns a large positive value. If they are identical, the coefficient, r_S = +1. If inversely related, it returns a large negative value, and if the rank of one variable decreases stepwise as the other variable increases

stepwise, $r_s = -1$. If the ranking of one variable is not related to that of the other, $r_s \approx 0$. (Kinnear & Gray 2006, p. 369, 377). The scatter plots for each association should also be viewed to ensure that relationships are linear because Spearman's Rank Correlation Coefficient will produce perfect correlation for certain non-linear relationships.

Factor analysis

Exploratory factor analysis attempts to simplify a correlation matrix of answers given to a series of questions (Kline 1994, pp. 7, 50) by determining how many underlying factors, linear correlations of variables weighted by the amount of the total variance that they explain (Kline 1994, p. 36). Factor analysis is ideal in this situation because "the data are complex and it is uncertain what the most important variables in the field are" (Kline 1994, p. 10), whilst requesting only easily provided information, as recommended by Busha & Harter (1980, p. 64). Factor analysis should prove much more informative than simpler methods such as summing the internal distances between answers, which often fail to differentiate between different answers in complex situations because different combinations of answers may produce the same overall scores (Oppenheim 1992, pp. 200-201).

Validity and reliability

The variables exploratory factor analysis are pre-defined, so the construct validity, as defined in Powell (1985, p. 36) is good because the choice of variables to be studied are well chosen and comprehensive. It is therefore essential that all variables that might be important in the topic are included, that the rationale for their inclusion is sound (Kline 1994, p. 72), and that the item pool is balanced as only if variables are measured can they contribute to possible underlying factors (Kline 1994, p. 12; Oppenheim 1992, p. 181). Kline (1994, p. 72) asserts that at least three variables are required for each factor, to distinguish between common and specific factors. Oppenheim (1992, p.147) points out that in order to establish whether any attitude question has been answered accurately,

several attitude questions asking similar things must be asked but it should be remembered that due to the sensitivity of wording effects, all three will be measuring similar but different aspects of the same attitude. This helps to stabilise strong attitude measurements and reduces the weight given to weaker attitudes but makes assessment of such measures' validity difficult (Oppenheim 1992, p. 147).

Factor analysis has poor internal validity, however, because even if the factors are rotated to true simple structure the interpretation of causal relationships is subjective and would require a confirmatory factor analysis to establish whether these factors were indeed present. If rotated to simple structure, the external and predictive validity of attitude scales (Sudman, Bradburn & Schwartz 1996, p. 110) and factor analysis (Kline 1994, p. 65) is good.

Sample size

The entire population was included in the survey. Principal Components Analysis was therefore the only suitable method because it measures factors within the sample without attempting extrapolation (Kline 1994, p. 49).

A large response was required because with sample sizes of less than 100, more than 20 times greater than the number of factors found after rotation to simple structure, and at least twice as large as the number of variables involved, are difficult to replicate reliably (Kline 1994, p. 73; Manly 1994, p. 105). The larger the sample size the better for factor analysis (Kline 1994, p. 73).

Since it consists of researchers of different levels of ability and experience, the sample is likely to be heterogeneous but provided the sample that responds is large enough to be representative, this will not matter. It might be argued that the professional and academic researchers should be considered separately as well as together to ensure the factors are the same for both groups but it is unlikely that the sample size will be large enough to facilitate this. (Kline 1994, p. 73)

If all the above criteria are fulfilled, *i.e.* at least 100 responses are returned that show no obvious signs of fatigue, acquiescence or attempted sabotage, then maximum likelihood factor analysis is best because its extrapolates the correlations to the underlying parent population (Kline 1994, p. 49) and indicates the number of statistically significant factors present (Kline 1994, p. 54).

Choice of method

Maximum likelihood factor analysis is powerful (Manly 1994, p. 105) but particularly sensitive to small sample size and samples that are not representative of the underlying population (Kline 1994, p. 49). If the sample size is small, principle components analysis (PCA) is more robust (Kline 1994, p. 49). PCA approximates to maximum likelihood analysis for large, representative samples when the test reliabilities, and therefore communalities, are high (Kline 1994, p. 49) but fails to separate out errors and therefore is less accurate than maximum likelihood factor analysis for large representative samples (Kline 1994, p. 50). The entire population is being sampled, however, and so extrapolation is not meaningful. PCA analyses the data set without attempting to extrapolate the data and is therefore robust whether analysing a minimal or comprehensive survey response, although larger samples are always preferred (Kline 1994, p. 73).

Number of significant factors

Either the number of factors present in the eigenvalue scree plot before the change of slope (Kline 1994 p. 75) or the number of factors with eigenvalues greater than unity (Manly 1994, p. 96), *i.e.* that explain more of the total variance than any one variable (Kline 1994, p. 38), may be used. Kline (1994, p. 75) recommended using the eigenvalue scree plot method to select the number of significant factors when the change in slope is clear and distinct, since many factors have eigenvalues greater than unity.

Factor rotation

In PCA, the first factor arrangement always shows a large, general factor with large positive loadings on each variable and bipolar factors with smaller positive and negative loadings for different variables, all of which are an artefact of the method and meaningless (Kline 1994, p. 39). Following Occam's razor, the factors must be rotated to give the simplest arrangement which explains the variation in the data, known as "simple structure" to ensure that the factors and their interpretation are clear and reproducible (Kline 1994, pp. 52, 64).

"Simple structure" was originally defined by (Thurstone 1947 *cited in* Kline 1994, p. 65) as meeting five criteria but modern analyses simply relies upon computers to resolve the factors to a simple structure where each factor has high loadings for a few variables and zero or near-zero loadings for all other variables (Kline 1994, p. 65).

Using the simplest methods first, the significant factors should first be rotated orthogonally because orthogonal rotations do not alter the proportion of variance explained by each factor (Kline 1994, p. 62). For this, the varimax rotation is most efficient (Kline 1994, p. 67; Manly 1994, p. 97), especially if the Kaiser normalisation is used (Manly 1994, p. 97). Assuming no correlation between factors, varimax maximises the sum of the squared loadings, producing loadings that are high or close to zero for each factor (Kline 1994, p. 68). If the factors are correlated with one another, orthogonal rotation will fail and oblique rotation will be required (Kline 1994, p. 67). Of the oblique methods, direct oblimin and maxplane have been shown by Hakstian (1971 *cited in* Kline 1994, p. 71) to be the most accurate, of which direct oblimin is the most reliable, although more than one method should always be used to check rotations because unusual cluster arrangements may lead to misleading arrangements using any method (Kline 1994, p. 71).

Identification of significant factors

Once simple structure has been obtained the variables loading each factor, will be resolved. If the sample size exceeds 100, factor loadings of magnitude of at least 0.3, which account for 9% of the total variance, may be considered significant (Kline 1994, p. 52). If the sample size is too small, the less sensitive approach proposed by Manly (1994, p. 101) of counting all loadings exceeding 0.5 in magnitude appears safer.

Corroboration of factor interpretations

The meaning of the different factors may be tentatively predicted from the size and sign of the different factor loadings on each variable but it is necessary to triangulate exploratory factor analysis data with other sources (Kline 1994, p. 180), which will have to be the subject of future research.

Choice of items to investigate

In the literature review, the authority and influence of authors, and the effect of place of publication were suggested as important factors affecting citation behaviour. Citation indices were therefore investigated. The perception of impact and the presence of positive findings were included to investigate the reported publication bias in scholarly literature towards positive findings (*e.g.* Torgerson 2006, p. 90) and to ascertain whether researchers view positive results as more relevant to future research. Funding was also included to determine whether researchers associated authority, author status, research impact and funding. Ghaebi (2003, p. 112) reported four factors were significantly related to researchers perception of document relevance, and so were also included.

Questionnaire theory

The more reliable the design of a questionnaire, the more reliable the data it will yield (Murray 1999, p. 148).

Why an online survey?

Use of a self-completed questionnaire enables inexpensive access to a much wider geographic sampling frame than would have been possible if the questions were to be completed through interviews especially given the inaccessibility of researchers (Busha & Harter 1980, p. 62) and facilitates the collection of a large volume of data in a short period of time, typically within 1-2 weeks (Powell 1985, p. 90). At least 100 responses are desired for the factor analysis (Kline 1994, p. 73), so questionnaires were the only feasible data collection method for this data.

It seemed reasonable to seek the corroborating qualitative data using the same instrument because this would at least reach a representative sample of researchers even if they did not answer the open-ended questions, and if participants answered, it was likely to encourage focus, brevity and clarity in their answers, making them easier to analyse. Since the online questionnaire was anonymous it was thought more likely to provide accurate answers that reflected researchers real views. Using a self-completed questionnaire also ensured that all participants had precisely the same stimulus and information and prevented variations in questions, ensuring reliability. (Powell 1985, p. 90; Fowler 2002, pp. 74, 81; Bourque & Fielder 2003, p. 9)

All the participants effectively received the questionnaire simultaneously, so the contextual influence of external events was standardised as much as possible (Bourque & Fielder 2003, p. 13).

The inability of participants to seek clarification or help completing the online questionnaire (Powell 1985, p. 91) meant that the questionnaire had to be completely self sufficient (Bourque & Fielder 2003, p. 7) and immediately

understandable to the target group. Unlike interviews, self-administered questionnaires are also unable to probe interesting answers further. Using familiar terms with well understood meanings was necessary to ensure the same response from each respondent (Fowler 2002, p. 81).

Sample frame

Professors and Readers are the most relevant researchers to the topic and may therefore be most motivated to answer it (Kelt 1996, p. 161), but they formed too small a sample frame from which to draw a sample of adequate size for the factor analysis to be replicable. The sample frame included all library and information science researchers in the UK, including all the library schools and any other researchers that may be found, otherwise it would have systematically excluded relevant individuals from the sample (Bourque & Fielder 2003, p. 15) and made it less representative and therefore less generalisable.

Nonresponse bias

More interested or highly opinionated researchers are more likely to be motivated to respond to a questionnaire than less opinionated sample members (Powell 1985, p. 91), posing the risk of a nonresponse bias (Busha & Harter 1980, p. 63). Whilst collecting more extreme views might make the qualitative analysis easier, the factors identified in the factor analysis will not be generalisable if the data received is not representative of the underlying population (Siegel & Castellan 1988, p. 34).

Online questionnaires

The anonymity that may be guaranteed with online questionnaires encourages frank and honest answers (Powell 1985, p. 90; Busha & Harter 1980, p. 62), especially for "sensitive" questions (Fink 1995, p. 17) and where egos are involved (Busha & Harter 1980, p.62), such as for the peer ranking question.

It is not known whether the inclusion of various progress indicators affects survey participation, although longer surveys and surveys containing many long answer open-ended questions have less participation (Bourque & Fielder 2003, p. 20). Most online surveys, including the one used, take 15-25 minutes to complete (Bourque & Fielder 2003, p. 107).

Pre-testing

Pre-testing was considered essential to check that the questions were understandable to the target group. Ideally, the sample for the pre-test should have been selected randomly from the same sample frame as will be used for the study (Powell 1985, p. 104) but this was not possible.

Questions should be as specific, concrete and balanced and easily understood as possible, and must be unidimensional, that is they must only ask about one thing at a time (*e.g.* Powell 1985, p. 100; Oppenheim 1992, p. 195; Janes 1999, p. 321). Clearly-worded, specific questions tend to involve less contextual information in the way that they are answered, and are therefore less affected by previous answers and emotional responses to earlier questions, making them more reliable (Sudman, Bradburn & Schwartz 1996, p. 84). Murray (1999, p. 149) suggests no question should be more than 20 words in length. Unclear, obscure, technical, slang terms should always be avoided, since respondents may not easily seek clarification (Busha & Harter 1980, p. 62) and wordage should be minimised (Busha & Harter 1980, p. 72). Studies have shown that asking whether events should be forbidden or allowed produce different results, so all questions should be phrased in a positive (allowing) manner (Powell 1985, p. 100). This also avoids the possibility of the word "not" being missed when participants read questions (Janes 1999, p. 324).

It appears unwise to include information or arguments in attitude scales (Schuman & Presser 1981, p. 185) for educated and relatively informed individuals such as researchers because it is impossible to include all pertinent facts and arguments and therefore the question may be biased, will certainly take much longer to answer, and if an argument is suggested that the researcher has

not thought of, may impress them and alter their opinion (Schuman & Presser 1981, p. 185).

Questions should be read repeatedly during pre-testing to uncover emotional charge and ambiguity of meaning (Busha & Harter 1980, p. 72).

Layout

Consistent symmetrical layout helps respondents to follow text (Bourque & Fielder 2003, p. 104). The questionnaire should be as short as possible and preferably with a pale blue background to ease reading (Kelt 1996, p. 161). Instructions to skip questions should be avoided where possible to simplify the questionnaire (Bourque & Fielder 2003, p. 33) and questions (Bourque & Fielder 2003, p. 105), possible answers and preferably relevant instructions should be on the same page, with tick boxes to the right of options. General web design principles such as visual flow, visual logic, the use of twelve point font, contrasting font and background colours and use of white space are relevant and important (Bourque & Fielder 2003, p. 110).

Order effects

Questions should be presented in a logical order from easy to more challenging questions (Bourque & Fielder 2003, p. 56) to increase interest and motivation as participants approach more searching questions but ensuring that the thoughts and information made available by one question does not influence the response to subsequent questions (Powell 1985, pp. 101). Moving from the general to the more specific is also desirable (Powell 1985, p. 103) because more specific questions are less vulnerable to context order effects (Sudman & Bradburn 1982, p. 143). Therefore the simple, Likert scale questions for the factor analysis were set first, followed by the more intrusive questions and finally the open-ended questions, which require most motivation to answer (Weingand 1993, p. 18).

As few questions were asked as possible to maximise participant response (Powell 1985, p. 104), avoid fatigue effects, and to ensure that all the data collected is easily analysable and relevant.

Attitudes are always context-dependent (Sudman, Bradburn & Schwartz 1996, p. 81). To avoid context/order effects, closed-ended answer questions were asked in a seemingly random order, but since the first question asked anchored the Likert scales for all subsequent questions (Sudman, Bradburn & Schwartz 1996, p. 96), it was chosen with care.

Questions on closely related topics suffer most from context effects (Schuman & Presser 1981, p. 27), perhaps because of the "halo effect", where participants may generalise their ratings from one answer to all similar answers in an attempt to make their responses consistent (Powell 1985, p. 99), especially when questions proceed from general to the more specific, although this may be helpful in guiding answers (Sudman & Bradburn 1982, p. 143). This is problematical, since multiple questions on similar topics are required for factor analysis (Kline 1994, p. 72). Such effects are assumed to be strongest with adjacent questions and decrease with distance (Sudman, Bradburn & Schwartz 1996, p. 120).

The effect is more pronounced if general questions follow specific questions (Sudman & Bradburn 1982, p. 143). It is therefore desirable to prevent participants from going backwards through the factor analysis questions to prevent them from changing previous answers (Bourque & Fielder 2003, p. 23).

Question type

Whether participants had conducted peer-assessments before and whether they had used citation indices to assess their peers was asked because past behaviour is a predictor of future behaviour (Powell 1985, p. 92). Direct questions about attitudes and behaviours were asked because the validity and reliability of indirect assessment methods are "open to question" (Powell 1985, p. 93).

Attitude questions were asked primarily because behavioural scientists have reached a consensus that although verbal expression does not predict subsequent action accurately, attitudinal questions may uncover predispositions that guide overt behaviour (Busha & Harter 1980, p. 67). Open-ended questions were included because it is necessary to corroborate the factors suggested by any exploratory factor analysis (Kline 1994, p. 181) and open questions are useful for exploratory research into complex issues with unknown dimensions, and where the research intends to explore processes and individuals' formulation of an issue (Sellitz *et al.* 1959, p. 262 *quoted in* Powell 1985, p, 92). Open-ended questions were necessary because it is impossible to provide an exhaustive list of all the possible pros and cons of peer review and find a means for participants to meaningfully rank them because of primacy and recency effects and because this might suggest reasons that the participants had not thought of, or participants might even use the answers provided to cover up a certain degree of ignorance (Powell 1985, p. 94) or more probably to avoid thinking hard about what is being asked.

Since open-ended questions "tend to discourage responses" because they take longer to answer (Powell 1985, p. 92), they were left until near the end of the questionnaire on the basis that whilst it would only be possible to answer a few complicated open-ended answers, as many responses as possible are needed for the factor analysis (Kline 1994, p. 73), so these were asked first. Answering the attitude questions might also make information available about peer review and citation metrics available in respondents minds, this context/order effect making it easier for them to answer the open-ended questions and provide deeper, more meaningful answers (Sudman, Bradburn & Schwartz 1996, p. 83).

Attitude scales

Thurstone scales are the most statistically rigorous attitude scales available, but there was no pool of judges available for pre-test calibration of the scales (Oppenheim 1992, p. 187) and so this was not used.

Powell (1985, p. 98) argued that Likert scales are only ordinal scales and cannot be regarded as interval scales, but their results strongly correlate with those of Thurstone scales (Oppenheim 1992, p. 195), so they were regarded as interval scales for the purposes of the factor analysis. Likert scales satisfy the requirements for factor analysis (Kline 1994). Providing the same answers for each question also avoids primacy and recency effects that affect long lists of variables, even in self-administered questionnaires (Sudman, Bradburn & Schwartz 1996, p. 123), and makes respondents think about their answers, further reducing context and order effects (Sudman, Bradburn & Schwartz 1996, p. 146). To further reduce primacy and recency effects, the advice of (Bourque & Fielder 2003, p. 99) to order answers vertically was ignored and the scale was displayed horizontally, so that answer options are read almost simultaneously.

Question answer formats were consistent to facilitate rapid, accurate completion of the questionnaire, although switching formats may prevent boredom effects (Powell 1985, p. 105). Since Likert scales may return increasing scores linearly in either direction along the scale, it is possible to reverse questions and reverse the weighting scale for answers (Oppenheim 1992, p. 195) to check for fatigue effects, acquiescence effects, and attempted sabotage.

Fatigue causes motivation to wane over time as questionnaire length increases and as participants progress through the questionnaire, combining with context/order effects to make responses to attitude and behavioural questions progressively less reliable (Sudman, Bradburn & Schwartz 1996, p. 154). Most respondents are believed to acquiesce to attitude questions, although this response is thought to decrease sharply with increasing education and should not be a problem with researchers who are used to critically appraisal (Schuman & Presser 1981, pp. 203-206).

The attitude scales should cover all possible grades of the attitude, contain equal numbers of positive and negative scale items (Oppenheim 1992, p. 181) and avoid any hint of positive or negative consequences of certain ratings that may cause assimilation effects (Sudman, Bradburn & Schwartz 1996, p. 102).

Reference to events, people or items that may engender an emotional response should also be avoided because such references have been shown to alter the question answered, as opinion of related issues surrounding the person or item are considered in the answer (Sudman, Bradburn & Schwartz 1996, p. 161). The questions seek to probe deeply guarded feelings, convictions and possibly irrational behaviours and anxieties that on introspection participants might even feel were unethical, so Oppenheim (1992, p. 179) recommends using familiar words such as trust, fear, respect and language related to feelings in attitude questions to probe statements generated by candid face-to-face interviews. In contrast, Busha & Harter (1980, p. 72) advise an impersonal approach that avoids rousing participants emotions, possibly because of the overarching order effects rousing emotions may have on all subsequent questions (Sudman, Bradburn & Schwartz 1996, p. 87).

Schuman & Presser (1981, p. 243) offer a way out of this by asking centrality questions that seek to know how important an issue is to the participant and behavioural questions that ask about predicted behaviours. Since behaviours are context-dependent and it is desirable to ask the same question throughout the factor analysis to make it easy and quick to complete, it is considered best to only ask centrality questions. This approach suffers from random measurement error most when opinions are weak and least when opinions are crystallised (Schuman & Presser 1981, p. 247) but since those with weak attitude strength tend to choose middle values (Sudman, Bradburn & Schwartz 1996, p. 126), this should not affect the factor analysis results. Acquiescence is also less pronounced with measures of centrality than with attitude scales requiring participants to agree or disagree with a statement (Fowler 2002, p. 94) but cannot overcome the tendency of some people to avoid giving extreme answers (Fowler 2002, p. 102).

Overall, it appears that questions that are as short and specific as possible, and that provide no background information are least likely to suffer from order or other effects.

Cover letter and ethical issues

The email requesting participation in the survey was sent from a University email address and branded (Fowler 2002, p. 148) with the University logo. The importance of the research and especially the need for their answers to the first section of the questionnaire, used for the factor analysis, was emphasised (Powell 1985, p. 106). The anonymity of answers, data protection and confidentiality was stated, together with assurances that respondents may skip questions that they did not wish to answer (Fowler 2002, p. 149). Fowler (2002, p. 149) also suggests that the cover letter emphasise the voluntary nature of the survey and that no negative consequences will result from refusing it.

The cover email promised participants in the survey a brief summary of the survey results, as suggested by Powell (1985, p. 106), to be constructed from the dissertation results and conclusions and emailed after the dissertation was completed. In order to facilitate follow-up emails to non-respondents, the cover email should also request that after completing the online questionnaire, the respondent should reply by return of email so that their name may be excluded from the repeat mailings.

The data collected was backed up, all copies held securely, and any printed material collected from printers immediately and stored securely. At the earliest possible opportunity, information was made anonymous.

Appendix C

Please consult CD-ROM (attached inside back cover).

Appendix D

Cover email

I am a Masters degree student at Loughborough University studying Information and Library Management. I am conducting research on how researchers decide which publications to cite, and their views on research quality in the subject area, under the supervision of Professor Charles Oppenheim. It is hoped that this research will illuminate an important but unexplored aspect of citation analysis and explore if and how citation analysis and traditional peer review differ.

You have been randomly selected from all the information and library management research groups in the UK. I would be grateful if you could take part in the survey stage of this research. If you are willing, please follow the hyperlink below and complete the online questionnaire. The questionnaire has been trialled and the short-answer questions should take no more than ten minutes to complete.

It is vitally important you complete at least the short first stage of the questionnaire in order for the conclusions to be reliable. All participants will be emailed a summary of the findings of the study. Your response and identity will be kept strictly confidential, and will not be attributed to you or your employees.

Please do not hesitate to contact Professor Oppenheim (C.Oppenheim@lboro.ac.uk) or myself (D.E.Bennett-06@student.lboro.ac.uk) if you have any queries or comments regarding this survey.

You may withdraw from this survey at any time and request the data that you submitted be destroyed.

Thank you.

Yours sincerely,
David E. Bennett.

Explanatory statement at end of questionnaire

Thank you for agreeing to participate in this research. Your answers are anonymous and will be treated in confidence.

This questionnaire comprises 34 questions.

The first 28 questions will be analysed together. There follows three open-ended questions assessing perceptions about citation analysis and peer review. The questionnaire ends with three more short-answer questions.

Appendix E

Most of the criteria have not been tested on LIS researchers and were examined to determine what other factors may influence citation behaviour. The examination of geographical publication bias was derived from the observations that UK/US publications are cited disproportionately often (*e.g.* Pasterkamp *et al.* 2007; Trimble 2005). Ghaebi (2003) used factor analysis to demonstrate that three further factors motivate senior Library and Information Science researchers to identify relevant publications for their research, which he described as "aboutness", "quality of information", and characteristics of information". Questions 1-4, 8, and 23-25 were adapted from Ghaebi (2003, p. 112). Text in italics indicates the factor being examined and was not included in the questionnaire.

Factor analysis

The following is a series of questions about factors influencing the citing of published research. How important are the following to YOU as reasons for citing research?

Strongly disagree	Slightly disagree	Unsure	Agree slightly	Agree strongly

Relevance

1. Works I cite are specific to my research need, for example to back up the argument that I am trying to make.
2. Works I cite are relevant to my research topic.
3. Works I cite cover the area of research comprehensively.
4. Works I cite cover the topic in depth.

Authors

5.   In general, I would prefer my published works to be cited by authors whose works I cite. [suggests they hold them in high esteem]

6.   Works I cite are mainly written by prestigious researchers.

7.   Works I cite are mainly written by less well known authors.

8.   I know the author(s) of the work.

9.   I regard the author(s) to be eminent experts in the field of research in which the work is published.

Rigour/quality of research

10.   Articles that I cite are published in peer-reviewed journals.

11.   Works that I cite comprehensively review the literature.

12.   I cite research that I consider to be rigorous.

Impact

13.   Works that I cite have been cited by other researchers.

14.   Works that I cite declare mainly positive findings.

15.   Works that I cite declare important findings.

Funding

16.   I keep aware of which areas within my field of research are being funded most.

17.   Works that I cite are mostly in well-funded areas of research.

18.   I ensure that my research is in a research area which is well funded.

Geographic effects of publication place

19.   Many of the articles that I cite are published in the UK or America.

20.   Many of the articles that I cite include articles published from European countries other than the UK

21.   Many of the articles that I cite are published in countries outside of the UK, EU and USA.

Quality of information

22.   I agree with the information in works that I cite.

23.   Works that I cite are consistent with what others have published.

24.   Researchers whom I greatly respect agree with the information in works that I cite.

25.   Works that I cite are clear and well written.

In the following questions, please indicate how well you agree with the following statements.

26.   In general, articles in more prestigious journals are cited more often than those published in less prestigious journals.

27.   In general, researchers with established research reputations are cited more often than less experienced researchers.

28.   In general, articles in areas where there is likely to be continued funding are cited more often.

Open questions

29.   What do you consider to be the strengths and weaknesses of peer assessment of researcher performance quality?

Citation indices are measures of the number of times that a researcher's publications have been cited in other published scholarly works.

30.   What do you consider to be the strengths and weaknesses of citation indices as measures of researcher performance quality?

31.   How do you feel peer review and citation indices compare as measures of researcher performance and potential?

Ranking analysis

32.   Please rank whom YOU consider to be the top five researchers **of your department** in decreasing order of overall research performance. Please include yourself if appropriate.

If you would prefer not to answer this question, please go onto the next question.

1.   _____ *(most accomplished researcher in dept)*
2.   _____
3.   _____
4.   _____
5.   _____ (*fifth most accomplished researcher in dept*)

33.   What position do you currently hold? *(please tick one box)*

Professor………………………………….. ☐
Reader………………………………….. ☐
Senior lecturer……………………....... ☐
Lecturer………………………………… ☐
Post-doctoral research office/associate ... ☐
Other…………………………………….☐

If other, please state:

_____

34.   If a citation index were discovered that correlated almost perfectly with peer review when compared using large random samples of researchers of all levels and abilities, would you consider using it for assessing researchers in your team(s)?

Almost certainly not	Probably not	Not sure either way	Probably	Almost certainly
☐	☐	☐	☐	☐

Order of the questions

To eliminate order effects, questions that might introduce ideas which might inform or alter responses to later questions will be asked after the questions that they might inform and more general questions in each topic will lead on to more precise questions as this minimises context/order effects (Sudman, Bradburn & Schwartz 1996, p. 84). Where possible, question ecology has been maintained to provide as logical a sense of flow as possible throughout the questionnaire.

The following orderings of questions were decided:

5 > (before) 6 > 7:   Question five is cryptic, and to ask question six before the others might prejudice the answers because it may be construed as more desirable to cite prestigious journals.

5-7 > 8-9:   Asking questions eight or nine first might induce participants to answer whether researchers known to them were considered prestigious.

10 > 11 > 12:   Questions were ordered from the general to the more specific, to minimise context/order effects.

13 > 14> 15:   Works that have important findings almost always have positive findings [positive literature bias and are cited more often, therefore if the questions are asked in any other order than that suggested here, the answer to the first question will practically dictate the answers to the other two.

19 > 20 > 21:   Disguising the purpose of the questions as much as possible.

Appendix F

Please consult CD-ROM (attached inside back cover).

Correlations of metrics calculated using *Web of Science* data

Spearman's rho			Hirsch's h-index score	Egghe's g-index score	Total citation count	Total publication count	Average number of citations per publication	Annual citation rate	h-range	g-range	h+ statistic	Lambda statistic (= g/h)	s statistic (=sum of citations to h/c_tot)	Median number of citations of cited publications	Median number of citations of all publications	v statistic (= h/total publication count)	Bihui's a statistic	Number of citations contributing to h
Hirsch's h-index score	Correlation Coefficient		1.000	.959**	.950**	.782**	.782**	.868**	.859**	.776**	.793**	.227	.287*	.750**	.467**	.329*	.859**	.947**
	Sig. (2-tailed)		.	.000	.000	.000	.000	.000	.000	.000	.000	.079	.012	.000	.000	.004	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
Egghe's g-index score	Correlation Coefficient		.959**	1.000	.987**	.780**	.823**	.912**	.873**	.898**	.895**	.582**	.294*	.804**	.444**	.295*	.946**	.986**
	Sig. (2-tailed)		.000	.	.000	.000	.000	.000	.000	.000	.000	.000	.010	.000	.000	.010	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
Total citation count	Correlation Coefficient		.950**	.987**	1.000	.790**	.815**	.916**	.864**	.884**	.912**	.544**	.258*	.813**	.439**	.264*	.958**	.992**
	Sig. (2-tailed)		.000	.000	.	.000	.000	.000	.000	.000	.000	.000	.025	.000	.000	.022	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
Total publication count	Correlation Coefficient		.782**	.780**	.790**	1.000	.381**	.659**	.758**	.698**	.675**	.293*	-.074	.523**	.040	-.220	.687**	.755**
	Sig. (2-tailed)		.000	.000	.000	.	.001	.000	.000	.000	.000	.022	.523	.000	.734	.058	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
Average number of citations per publication	Correlation Coefficient		.782**	.823**	.815**	.381**	1.000	.832**	.660**	.720**	.794**	.454**	.536**	.794**	.744**	.664**	.854**	.850**
	Sig. (2-tailed)		.000	.000	.000	.001	.	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
Annual citation rate	Correlation Coefficient		.868**	.912**	.916**	.659**	.832**	1.000	.752**	.823**	.821**	.527**	.256*	.767**	.504**	.334**	.876**	.913**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.	.000	.000	.000	.000	.028	.000	.000	.004	.000	.000
	N		74	74	74	74	74	74	74	74	74	59	74	74	74	73	74	74
h-range	Correlation Coefficient		.859**	.873**	.864**	.758**	.660**	.752**	1.000	.870**	.828**	.456**	.079	.645**	.361**	.096	.816**	.862**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.	.000	.000	.000	.496	.000	.001	.413	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
g-range	Correlation Coefficient		.776**	.898**	.884**	.698**	.720**	.823**	.870**	1.000	.908**	.777**	.156	.736**	.334**	.129	.903**	.884**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.	.000	.000	.179	.000	.003	.271	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
h+ statistic	Correlation Coefficient		.793**	.895**	.912**	.675**	.794**	.821**	.828**	.908**	1.000	.709**	.309**	.827**	.406**	.225	.973**	.927**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.	.000	.007	.000	.000	.052	.000	.000
	N		76	76	76	76	76	74	76	76	76	61	76	76	76	75	76	76
Lambda statistic (= g/h)	Correlation Coefficient		.227	.582**	.544**	.293*	.454**	.527**	.456**	.777**	.709**	1.000	-.085	.456**	.055	-.212	.714**	.549**
	Sig. (2-tailed)		.079	.000	.000	.022	.000	.000	.000	.000	.000	.	.517	.000	.676	.101	.000	.000
	N		61	61	61	61	61	59	.870	61	61	61	61	61	61	61	61	61
s statistic (=sum of citations to h/c_tot)	Correlation Coefficient		76*	.294*	.258*	-.074	.536**	.256*	.000	.156	.309**	-.085	1.000	.512**	.392**	.772**	.390**	.329**
	Sig. (2-tailed)		.776	.010	.025	.523	.000	.028	76	.179	.007	.517	.	.000	.000	.000	.001	.004
	N		.000	76	76	76	74	76	76	76	61	76	76	76	75	76	76
Median number of citations of cited publications	Correlation Coefficient		76**	.804**	.813**	.523**	.794**	.767**	.000**	.736**	.827**	.456**	.512**	1.000	.434**	.464**	.875**	.846**
	Sig. (2-tailed)		.793	.000	.000	.000	.000	.000	76	.000	.000	.000	.000	.	.000	.000	.000	.000
	N		.000	76	76	76	76	74	.456	76	76	61	76	76	76	75	76	76
Median number of citations of all publications	Correlation Coefficient		76**	.444**	.439**	.040	.744**	.504**	.000**	.334**	.406**	.055	.392**	.434**	1.000	.697**	.446**	.454**
	Sig. (2-tailed)		.227	.000	.000	.734	.000	.000	61	.003	.000	.676	.000	.000	.	.000	.000	.000
	N		.079	76	76	76	76	74	.079	76	76	61	76	76	76	75	76	76
v statistic (= h/total publication count)	Correlation Coefficient		61**	.295*	.264*	-.220	.664**	.334**	.496	.129	.225	-.212	.772**	.464**	.697**	1.000	.321**	.313**
	Sig. (2-tailed)		.287	.010	.022	.058	.000	.004	76	.271	.052	.101	.000	.000	.000	.	.005	.006
	N		.012	75	75	75	75	73	.645	75	75	61	75	75	75	75	75	75
Bihui's a statistic	Correlation Coefficient		76**	.946**	.958**	.687**	.854**	.876**	.000**	.903**	.973**	.714**	.390**	.875**	.446**	.321**	1.000	.972**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	76	.736	.000	.000	.001	.000	.000	.005	.	.000
	N		76	76	76	76	76	74	.361	76	76	61	76	76	76	75	76	76
Number of citations contributing to h	Correlation Coefficient		.947**	.986**	.992**	.755**	.850**	.913**	.001**	76*	.927**	.549**	.329**	.846**	.454**	.313**	.972**	1.000
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	76	.334	.000	.000	.004	.000	.000	.006	.000	.
	N		76	76	76	76	76	74	.003	76	61	76	76	76	75	76	76

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations of metrics calculated using *ScHolar Index* data

		Hirsch's h-index score	Egghe's g-index score	Total citation count	Total publication count	Average number of citations per publication	h-range	g-range	h+ statistic	Lambda statistic (= g/h)	s statistic (=sum of citations to h/c_tot)	Median number of citations of cited publications	Median number of citations of all publications	v statistic (= h/total publication count)	Bihui's a statistic	Number of citations contributing to h
Spearman's rho Hirsch's h-index score	Correlation Coefficient	1.000	.965**	.971**	.892**	.785**	.880**	.827**	.862**	.038	.032	.552**	.655**	.154	.842**	.951**
	Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000	.745	.767	.000	.000	.156	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Egghe's g-index score	Correlation Coefficient	.965**	1.000	.988**	.875**	.836**	.921**	.922**	.933**	.317**	.078	.577**	.620**	.153	.933**	.985**
	Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000	.005	.474	.000	.000	.161	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Total citation count	Correlation Coefficient	.971**	.988**	1.000	.887**	.829**	.913**	.898**	.923**	.249*	.028	.577**	.658**	.128	.917**	.984**
	Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000	.029	.797	.000	.000	.240	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Total publication count	Correlation Coefficient	.892**	.875**	.887**	1.000	.525**	.792**	.736**	.739**	.089	-.171	.572**	.437**	-.218*	.720**	.835**
	Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000	.000	.441	.116	.711	.000	.043	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Average number of citations per publication	Correlation Coefficient	.785**	.836**	.829**	.525**	1.000	.790**	.835**	.869**	.330**	.337**	.345**	.766**	.541**	.883**	.869**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.003	.002	.001	.000	.000	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
h-range	Correlation Coefficient	.880**	.921**	.913**	.792**	.790**	1.000	.936**	.906**	.302**	.075	.711**	.595**	.142	.876**	.930**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000	.000	.008	.493	.711	.000	.192	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
g-range	Correlation Coefficient	.827**	.922**	.898**	.736**	.835**	.936**	1.000	.933**	.520**	.168	.711**	.558**	.190	.935**	.926**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.	.000	.000	.122	.000	.000	.079	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
h+ statistic	Correlation Coefficient	.862**	.933**	.923**	.739**	.869**	.906**	.933**	1.000	.414**	.214*	.630**	.586**	.225*	.976**	.967**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.	.000	.048	.000	.000	.037	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Lambda statistic (= g/h)	Correlation Coefficient	.038	.317**	.249*	.089	.330**	.302**	.520**	.414**	1.000	.088	.197	-.071	-.055	.502**	.296**
	Sig. (2-tailed)	.745	.005	.029	.441	.003	.008	.000	.000	.	.446	.086	.541	.637	.000	.009
	N	77	77	77	77	77	77	77	77	77	77	77	77	77	77	77
s statistic (=sum of citations to h/c_tot)	Correlation Coefficient	.032	.078	.028	-.171	.337**	.075	.168	.214*	.088	1.000	.469**	.015	.691**	.278**	.128
	Sig. (2-tailed)	.767	.474	.797	.116	.002	.493	.122	.048	.446	.	.000	.893	.000	.010	.239
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Median number of citations of cited publications	Correlation Coefficient	.552**	.577**	.572**	.345**	.711**	.508**	.549**	.630**	.197	.469**	1.000	.456**	.556**	.636**	.613**
	Sig. (2-tailed)	.000	.000	.000	.001	.000	.000	.000	.000	.086	.000	.	.000	.000	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Median number of citations of all publications	Correlation Coefficient	.655**	.620**	.658**	.437**	.766**	.595**	.558**	.586**	-.071	.015	.456**	1.000	.422**	.563**	.642**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.541	.893	.000	.	.000	.000	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
v statistic (= h/total publication count)	Correlation Coefficient	.154	.153	.128	-.218*	.541**	.142	.190	.225*	-.055	.691**	.556**	.422**	1.000	.260*	.195
	Sig. (2-tailed)	.156	.161	.240	.043	.000	.192	.079	.037	.637	.000	.000	.000	.	.015	.072
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Bihui's a statistic	Correlation Coefficient	.842**	.933**	.917**	.720**	.883**	.876**	.935**	.976**	.502**	.278**	.636**	.563**	.260*	1.000	.957**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.010	.000	.000	.015	.	.000
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86
Number of citations contributing to h	Correlation Coefficient	.951**	.985**	.984**	.835**	.869**	.930**	.926**	.967**	.296**	.128	.613**	.642**	.195	.957**	1.000
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.009	.239	.000	.000	.072	.000	.
	N	86	86	86	86	86	86	86	86	77	86	86	86	86	86	86

. Correlation is significant at the 0.01 level (2-tailed).

*.** Correlation is significant at the 0.05 level (2-tailed).

Appendix H

Web of Science		results					
h-index							
Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
34810.5	75	464.14	Gender	3937.435	8.483291	0.003584	0.996416
			RAE	665.5469	1.433936	0.838276	0.161724
			Interaction	5438.247	11.71683	0.019586	0.980414
h-range							
Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
30788.5	75	410.5133	Gender	3565.613	8.685743	0.003207	0.996793
			RAE	529.8908	1.290801	0.862934	0.137066
			Interaction	3035.902	7.395381	0.116412	0.883588
h+							
Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
35171.5	75	468.9533	Gender	3567.662	7.607711	0.005812	0.994188
			RAE	249.0254	0.531024	0.970411	**0.029589**
			Interaction	1553.851	3.313444	0.506805	0.493195
v							
Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
34779	74	469.9865	Gender	319.9694	0.680806	0.409309	0.590691
			RAE	1767.928	3.761656	0.439222	0.560778
			Interaction	2265.096	4.819491	0.306325	0.693675
Λ							
Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
17167	60	286.1167	Gender	126.5515	0.442307	0.506011	0.493989
			RAE	1372.993	4.798716	0.308581	0.691419
			Interaction	1902.375	6.64895	0.155644	0.844356

ScHolar Index results

h-index

Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
51780.5	85	609.1824	Gender	327.4773	0.537568592	0.463442	0.536558
			RAE	1176.485	1.931252456	0.748402	0.251598
			Interaction	2405.551	3.948820012	0.412977	0.587023

h-range

Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
49968	85	587.8588	Gender	360.3197	0.612935808	0.433685	0.566315
			RAE	1525.419	2.594873061	0.627732	0.372268
			Interaction	2369.738	4.031134611	0.401809	0.598191

h+

Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
52765	85	620.7647	Gender	389.3788	0.627256652	0.428363	0.571637
			RAE	1782.851	2.872024272	0.579465	0.420535
			Interaction	2362.653	3.806036148	0.432892	0.567108

h+

Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
52859.5	85	621.8765	Gender	13.54369	0.021778742	0.882677	0.117323
			RAE	878.2864	1.41231648	0.842052	0.157948
			Interaction	2320.716	3.731795091	0.443518	0.556482

Λ

Total SS	Total df	Total MS	Factor	SS	SS/MS	Chi-sq	P
37400.5	76	492.1118	Gender	45.70198	0.0928691	0.760561	0.239439
			RAE	1799.846	3.657392836	0.454353	0.545647
			Interaction	437.9218	0.889882692	0.926008	0.073992

Correlation matrix of metrics calculated using Web of Science and ScHolar Index

			Hirsch's h-index score	h-range	h+	v statistic (= h/total publication count)	g	Lambda statistic (= g/h)	Total number of publications contributing to h listed by Web of Science	Total number of citaitons contributing to h listed by Web of Science	Hirsch's h-index score	h-range	Scholar_h_plus	v statistic (= h/total publication count)	g	Lambda statistic (= g/h)	Total number of publications contributing to h listed by ScHolar Index	Total number of citations contributing to h listed by ScHolar Index
Spearman's rho	Hirsch's h-index score	Correlation Coefficient	1.000	.858**	.804**	.379**	.963**	.650**	.793**	.956**	.531**	.456**	.544**	-.002	.553**	.187	.473**	.535**
		Sig. (2-tailed)	.	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.988	.000	.117	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	h-range	Correlation Coefficient	.858**	1.000	.848**	.142	.881**	.631**	.763**	.871**	.559**	.526**	.556**	.110	.584**	.126	.443**	.555**
		Sig. (2-tailed)	.000	.	.000	.236	.000	.000	.000	.000	.000	.000	.000	.360	.000	.295	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	h+	Correlation Coefficient	.804**	.848**	1.000	.270*	.900**	.832**	.687**	.915**	.411**	.391**	.439**	.041	.454**	.141	.363**	.415**
		Sig. (2-tailed)	.000	.000	.	.023	.000	.000	.000	.000	.000	.001	.000	.734	.000	.240	.002	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	v statistic (= h/total publication count)	Correlation Coefficient	.379**	.142	.270*	1.000	.366**	.450**	-.125	.335**	-.064	-.049	.023	.172	-.017	.210	-.111	-.049
		Sig. (2-tailed)	.001	.236	.023	.	.002	.000	.300	.004	.595	.687	.848	.153	.888	.079	.357	.686
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	g	Correlation Coefficient	.963**	.881**	.900**	.366**	1.000	.810**	.778**	.988**	.521**	.460**	.549**	.041	.558**	.219	.453**	.526**
		Sig. (2-tailed)	.000	.000	.000	.002	.	.000	.000	.000	.000	.000	.000	.735	.000	.066	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Lambda statistic (= g/h)	Correlation Coefficient	.650**	.631**	.832**	.450**	.810**	1.000	.495**	.788**	.263*	.267*	.343**	.146	.330**	.273*	.209	.279*
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000	.000	.027	.024	.003	.224	.005	.021	.080	.019
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Total number of publications contributing to h listed by Web of Science	Correlation Coefficient	.793**	.763**	.687**	-.125	.778**	.495**	1.000	.791**	.562**	.488**	.549**	-.038	.563**	.093	.532**	.557**
		Sig. (2-tailed)	.000	.000	.000	.300	.000	.000	.	.000	.000	.000	.000	.756	.000	.442	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Total number of citaitons contributing to h listed by Web of Science	Correlation Coefficient	.956**	.871**	.915**	.335**	.988**	.788**	.791**	1.000	.510**	.447**	.527**	.006	.540**	.188	.457**	.512**
		Sig. (2-tailed)	.000	.000	.000	.004	.000	.000	.000	.	.000	.000	.000	.960	.000	.116	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Hirsch's h-index score	Correlation Coefficient	.531**	.559**	.411**	-.064	.521**	.263*	.562**	.510**	1.000	.858**	.851**	.060	.963**	.171	.890**	.972**
		Sig. (2-tailed)	.000	.000	.000	.595	.000	.027	.000	.000	.	.000	.000	.621	.000	.154	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	h-range	Correlation Coefficient	.456**	.526**	.391**	-.049	.460**	.267*	.488**	.447**	.858**	1.000	.912**	.182	.914**	.380**	.720**	.900**
		Sig. (2-tailed)	.000	.000	.001	.687	.000	.024	.000	.000	.000	.	.000	.129	.000	.001	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Scholar_h_plus	Correlation Coefficient	.544**	.556**	.439**	.023	.549**	.343**	.549**	.527**	.851**	.912**	1.000	.220	.925**	.477**	.686**	.904**
		Sig. (2-tailed)	.000	.000	.000	.848	.000	.003	.000	.000	.000	.000	.	.066	.000	.000	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	v statistic (= h/total publication count)	Correlation Coefficient	-.002	.110	.041	.172	.041	.146	-.038	.006	.060	.182	.220	1.000	.105	.253*	-.334**	.051
		Sig. (2-tailed)	.988	.360	.734	.153	.735	.224	.756	.960	.621	.129	.066	.	.383	.034	.004	.675
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	g	Correlation Coefficient	.553**	.584**	.454**	-.017	.558**	.330**	.563**	.540**	.963**	.914**	.925**	.105	1.000	.382**	.846**	.983**
		Sig. (2-tailed)	.000	.000	.000	.888	.000	.005	.000	.000	.000	.000	.000	.383	.	.001	.000	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Lambda statistic (= g/h)	Correlation Coefficient	.187	.126	.141	.210	.219	.273*	.093	.188	.171	.380**	.477**	.253*	.382**	1.000	.131	.311**
		Sig. (2-tailed)	.117	.295	.240	.079	.066	.021	.442	.116	.154	.001	.000	.034	.001	.	.276	.008
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Total number of publications contributing to h listed by ScHolar Index	Correlation Coefficient	.473**	.443**	.363**	-.111	.453**	.209	.532**	.457**	.890**	.720**	.686**	-.334**	.846**	.131	1.000	.876**
		Sig. (2-tailed)	.000	.000	.002	.357	.000	.080	.000	.000	.000	.000	.000	.004	.000	.276	.	.000
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71
	Total number of citations contributing to h listed by ScHolar Index	Correlation Coefficient	.535**	.555**	.415**	-.049	.526**	.279*	.557**	.512**	.972**	.900**	.904**	.051	.983**	.311**	.876**	1.000
		Sig. (2-tailed)	.000	.000	.000	.686	.000	.019	.000	.000	.000	.000	.000	.675	.000	.008	.000	.
		N	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71	71

Web of Science rows (rows 1–8) · *ScHolar Index rows* (rows 9–16)

Web of Science columns (columns 1–8) · *ScHolar Index* columns (columns 9–16)

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix J

Departments often comprise researchers working in different subfields. A proposed method of comparing subfields without normalisation, is to subtract the median metric score for the relevant subfield from that of assessed individuals.

For individuals whose research spanned several subfields, the weighted mean (weighted for the number of publications in each subfield) of the medians of the subfields in which they published could be subtracted from individuals' metric scores. The distances between the individuals in a research group, department or institution and their respective subfield medians could then be summed to give a score for the entire department. Random differences should cancel one another out to yield a reliable departmental score.

Difficulties arise when the distribution of researchers working in different subfields is uneven. In psychology, where departments often comprise the entire population of researchers in a subfield (Case & Higgins 2000, p. 642), researchers may be readily compared with their own subfield average but not with other departments or institutions. It is unclear how this could be resolved, since even after effective subfield normalisation, comparisons between subfields assume both fields being compared are in a steady state and are not increasing or decreasing in activity.

Appendix K

Correlations of *g* with other metrics

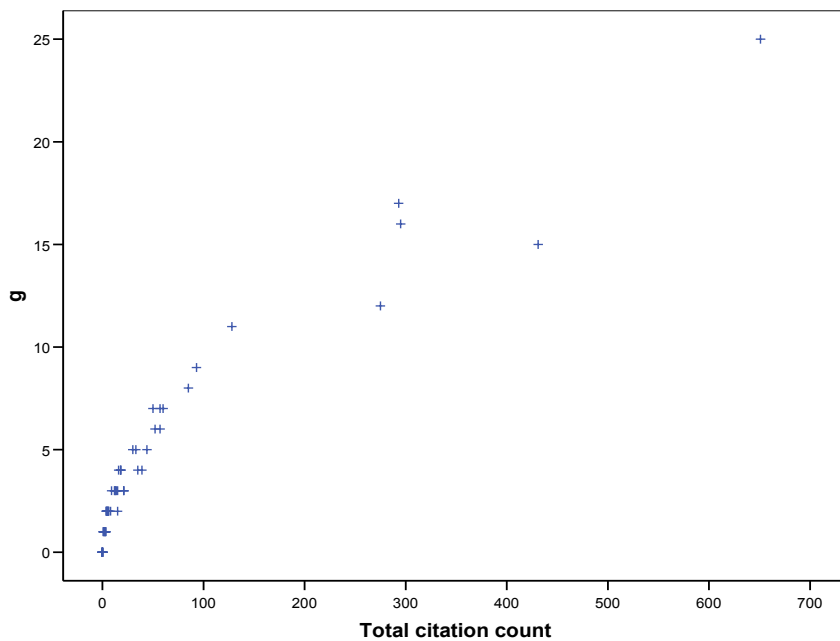The following scatter plots were derived from *Web of Science* data only.

Figure K1. Scatter plot of *h* against total citation counts

Figure K1 suggests a curvilinear relationship exists between *g* and total citation counts, as predicted from the relationship with *h*.
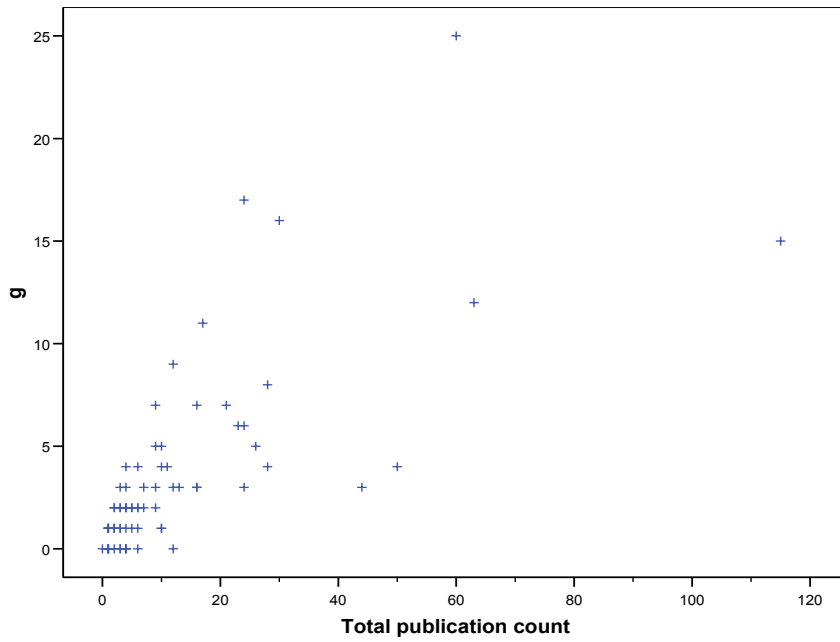
Figure K2. Scatter plot of *g* against total publication counts
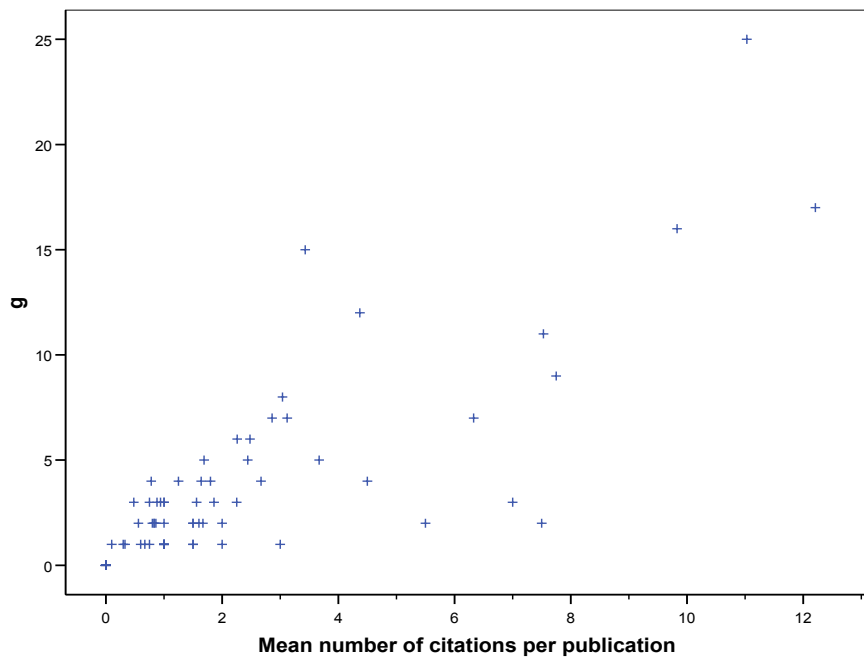

Figure K3. Scatter plot of *h* against mean numbers of citations per publication

Figures K2 and K3 showed a positive association with increasing scatter with both increasing *g* and increasing total publication count, respectively.
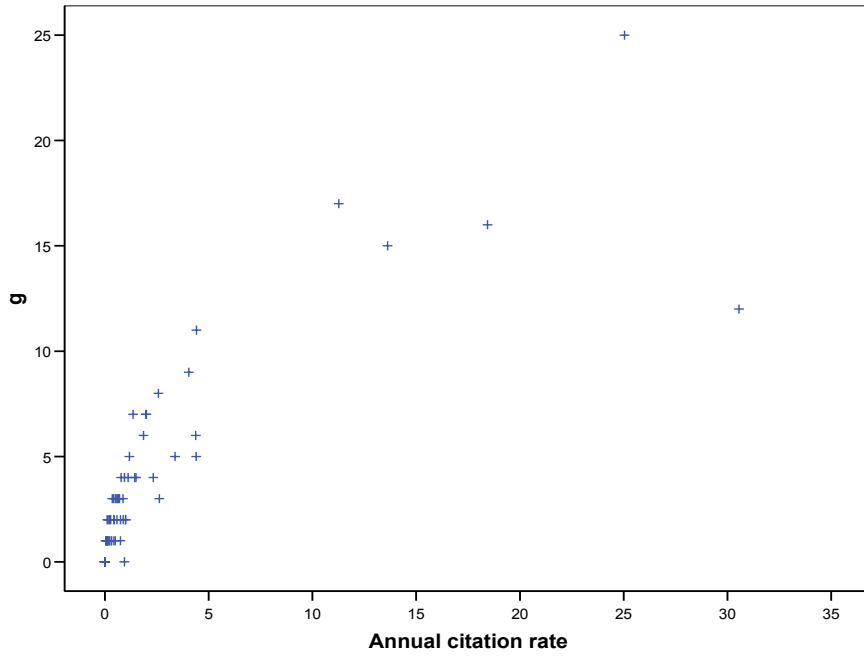
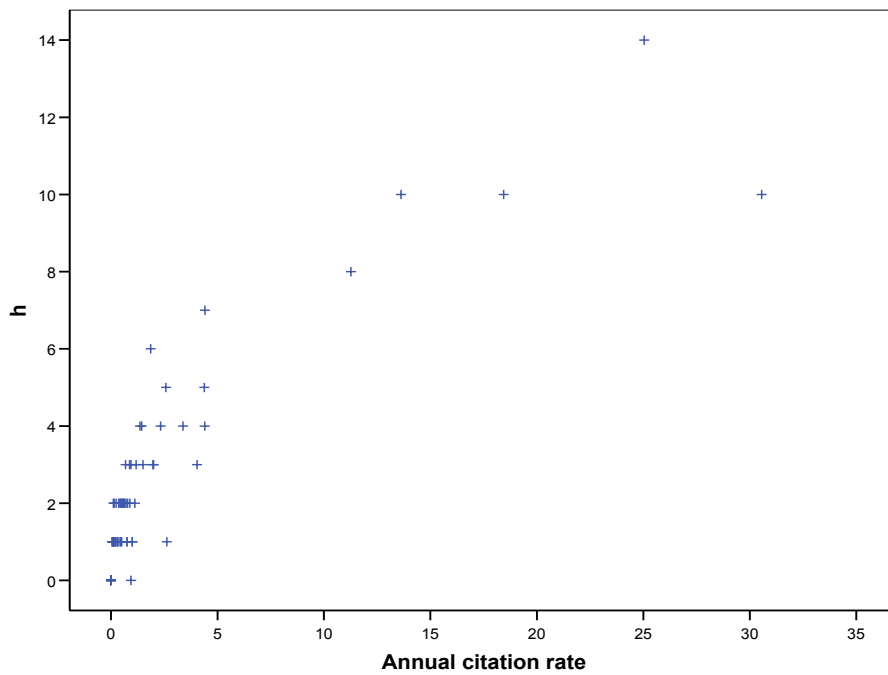**Figure K4.** Scatter plot of *h* against total citation counts

Figure K5. *h* against annual citation rates

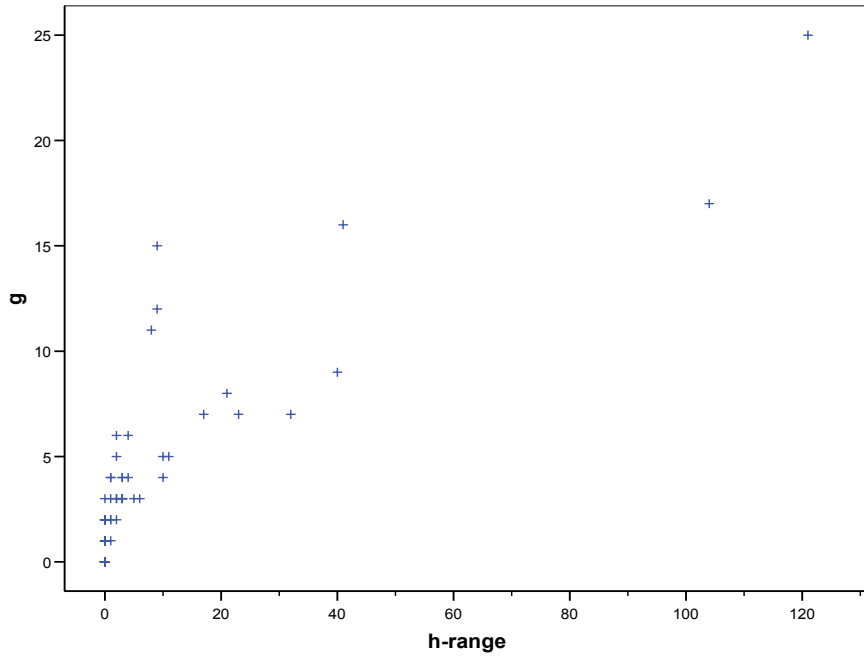Figures K4 and K5 appeared to show curvilinear association of *h* and *g* with annual citation rates.

Figure K6. Scatter plot of *g* against *h*-range
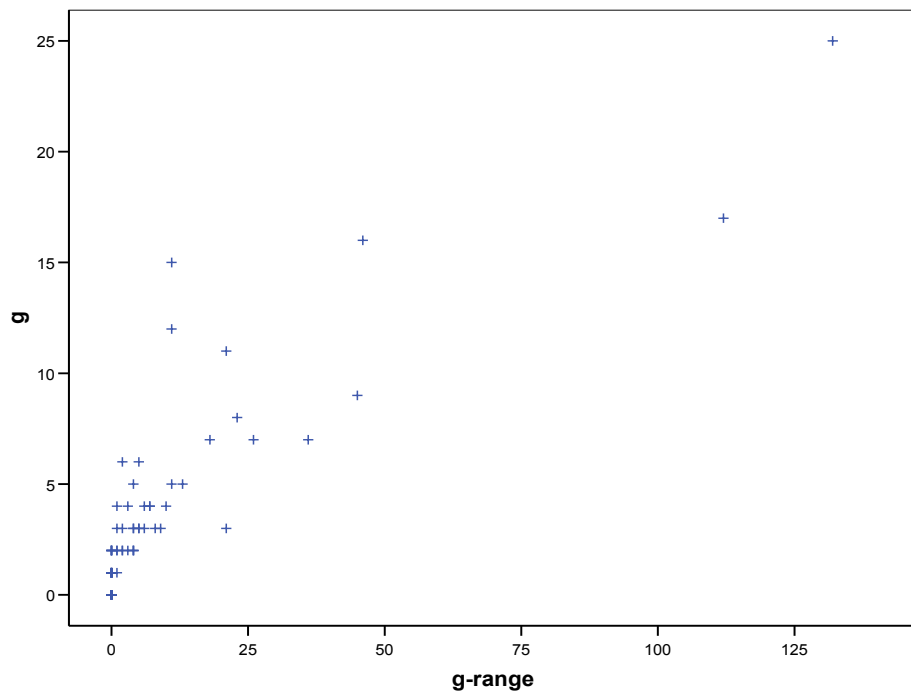

Figure K6. Scatter plot of *g* against *g*-range

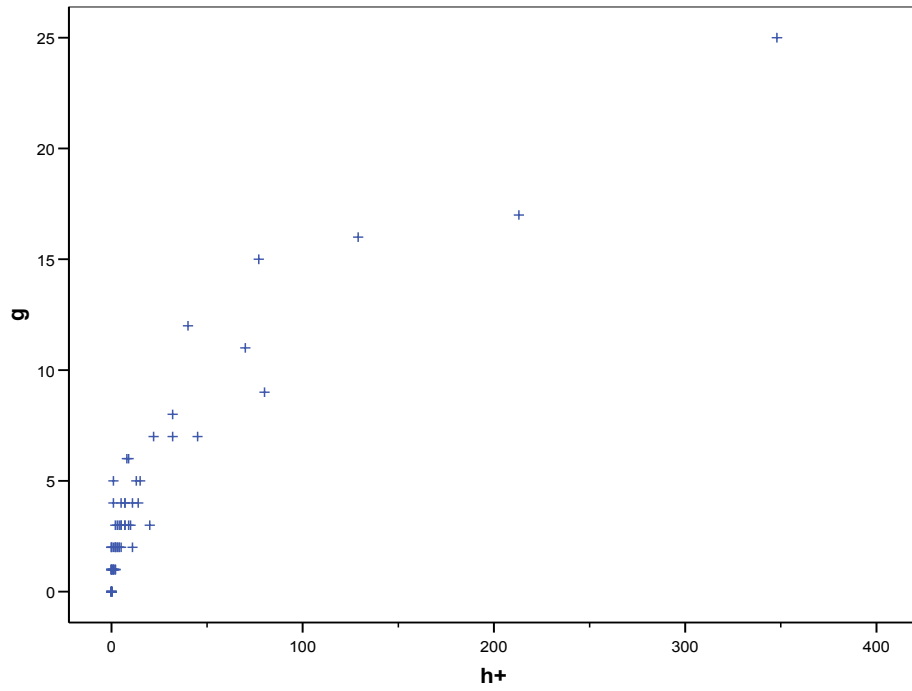Figure K5 and K6 show *g* was linearly associated with *h*-range and *g*-range.

171

Figure K7. Scatter plot of *g* against h^+ score*s*

Figure K7 showed a distinct curvilinear association between *g*-index and h^+ score*s*.
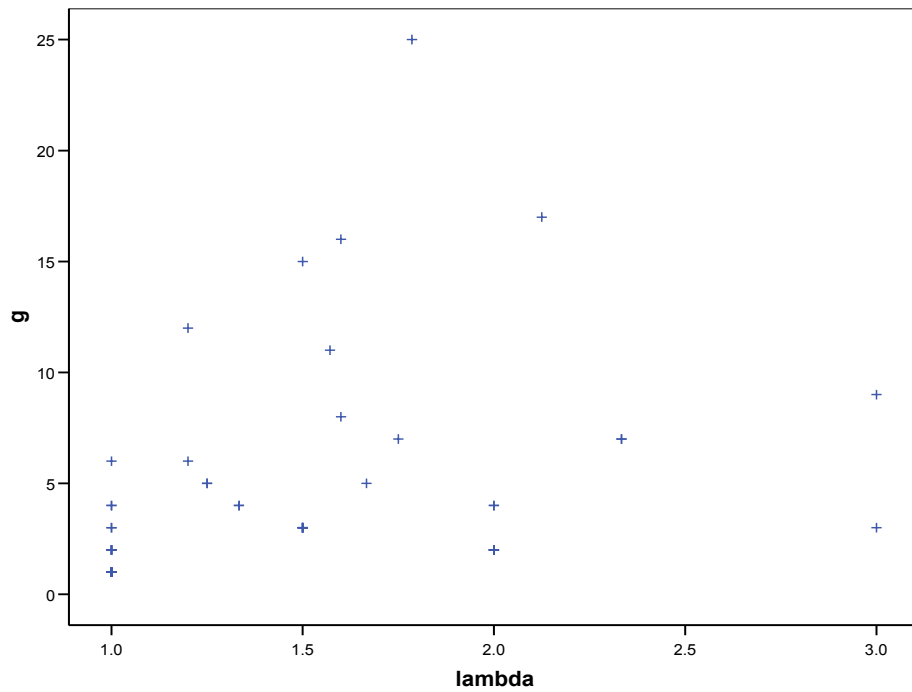
Figure K8. Scatter plot of *g* against lambda
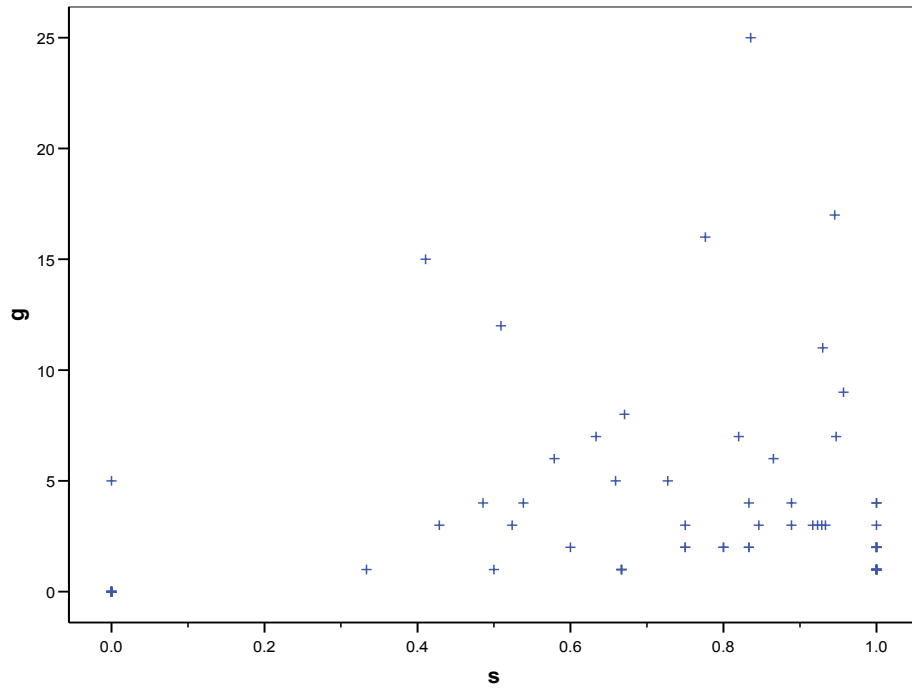
Figure K9. Scatter plot of *g* against *s* score*s*
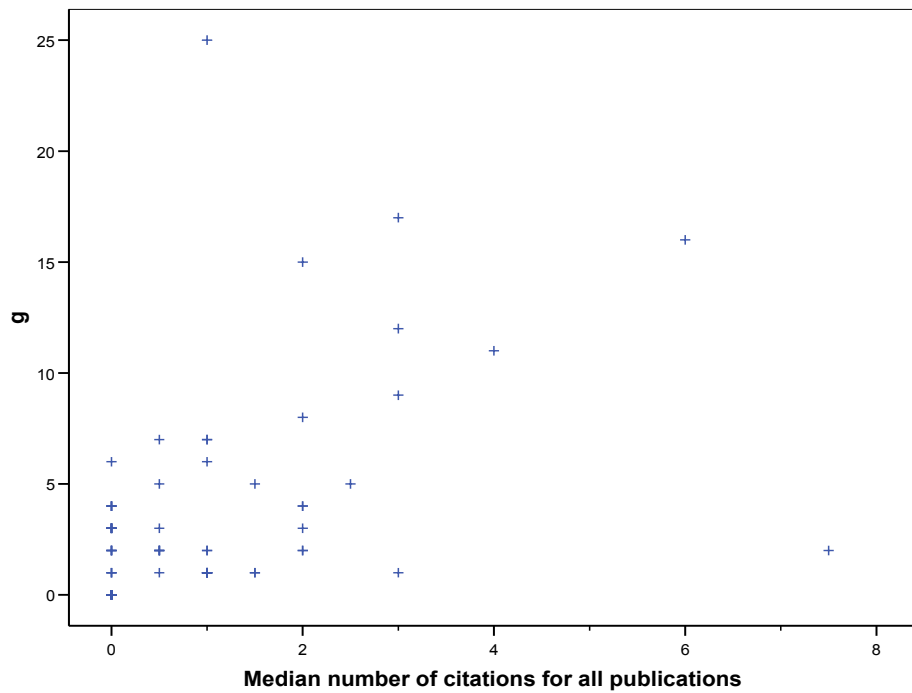
Figure K10. Scatter plot of *g* median number of citations for all publications

173
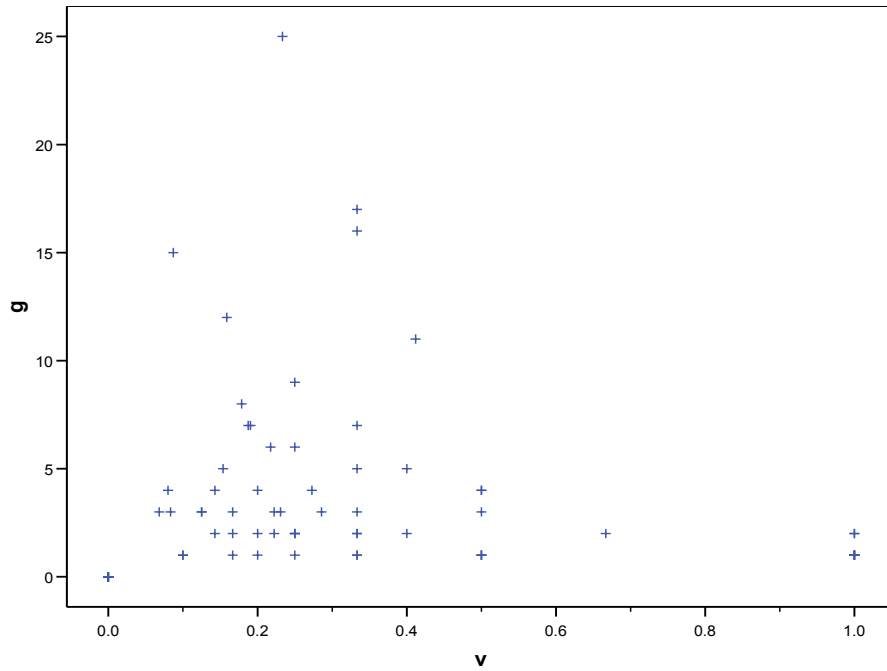
Figure K11. Scatter plot of *g v* score*s*

Figures K9 and K11 show no clear associations between *g* and *s* and *v*, respectively. The association shown in figure K10 between *g* and median citation counts was clear, positive and linear.
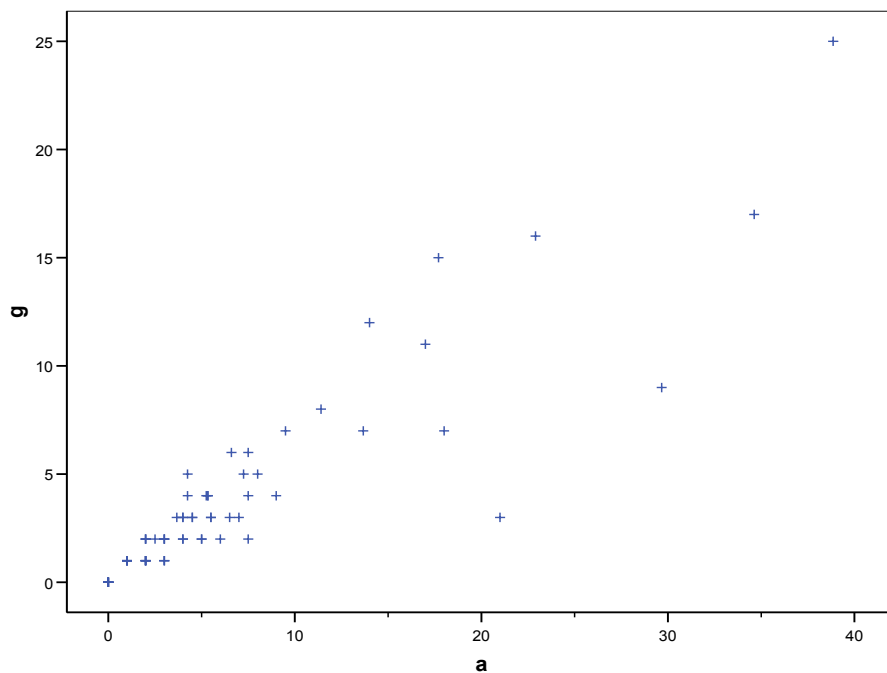
Figure K12. Scatter plot of *g* against a

Figure K12 shows *g* was strongly and linearly positively correlated with *a*.
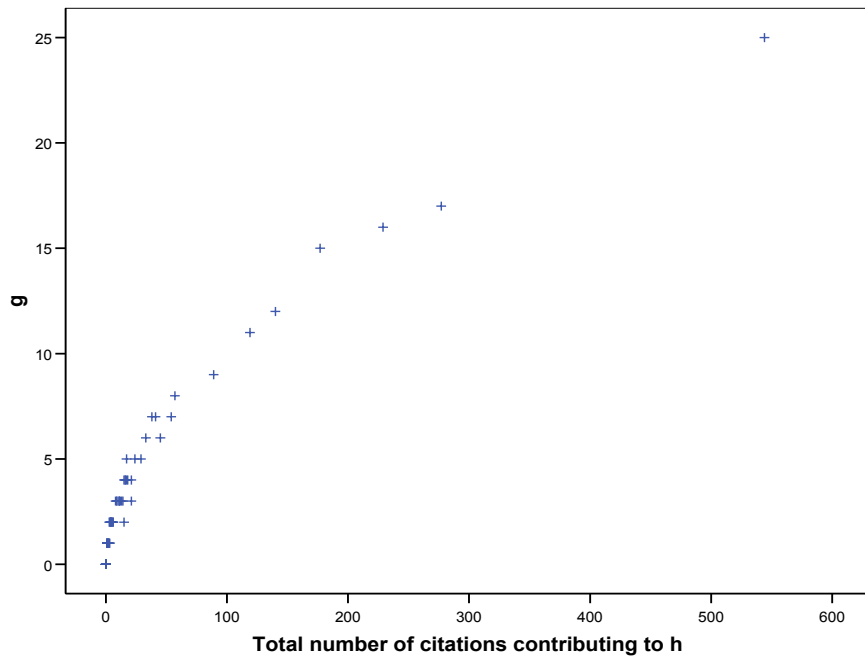
Figure K13. Scatter plot of total numbers of citations contributing to *h*

Figure K13 shows that *g* was correlated curvilinearly with the total number of citations contributing to *h*.
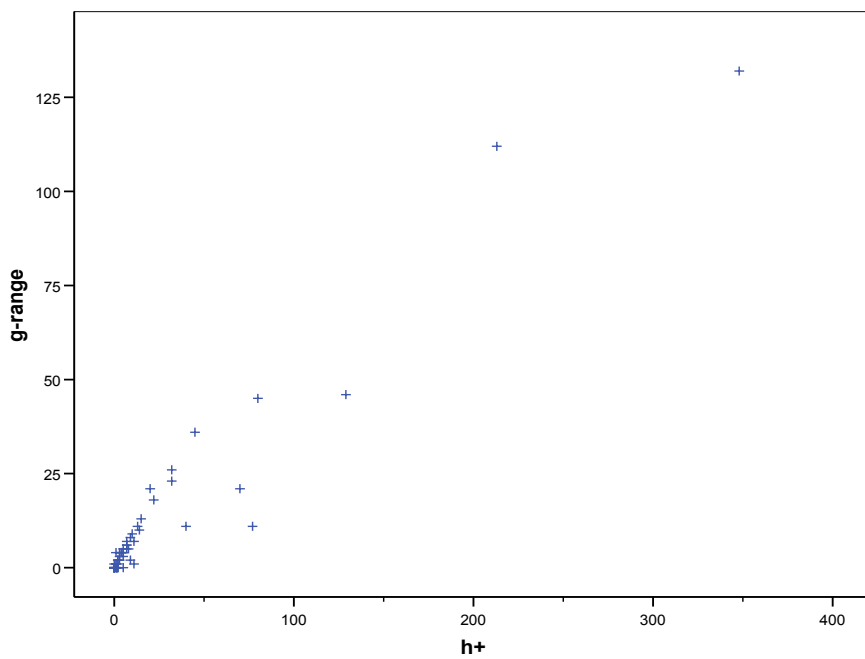
Figure K14. Scatter plot of *g* against h^+

Figure K14 shows that the linear relationship between g-ranges was slightly more reliable than that between h-ranges and h^+.
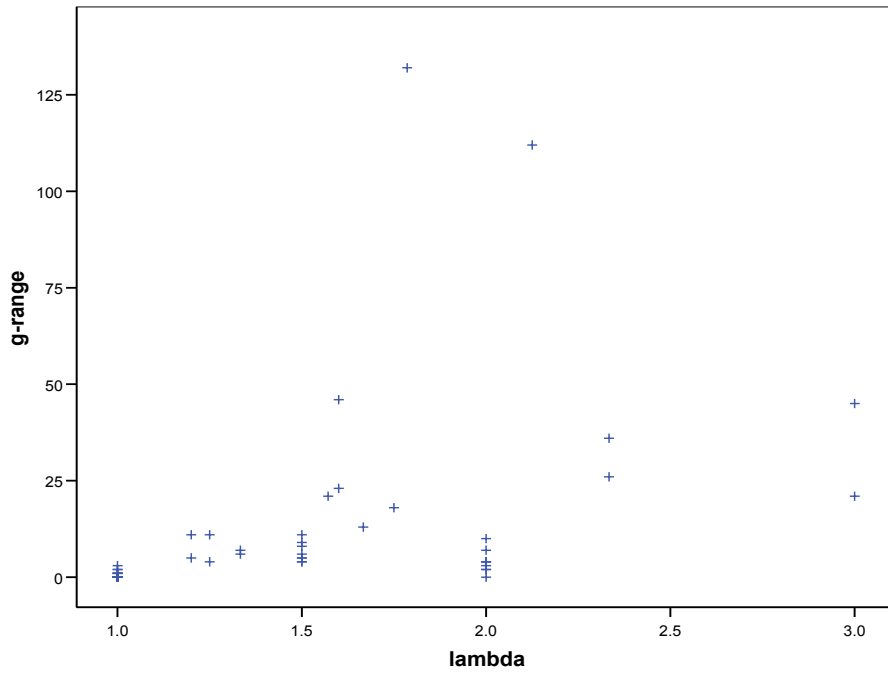
Figure K15. Scatter plot of *g*-ranges against lambda score*s*

Figure K15 showed that although strongly correlated, *g*-range increases very little with Λ, suggesting that the two are not associated.
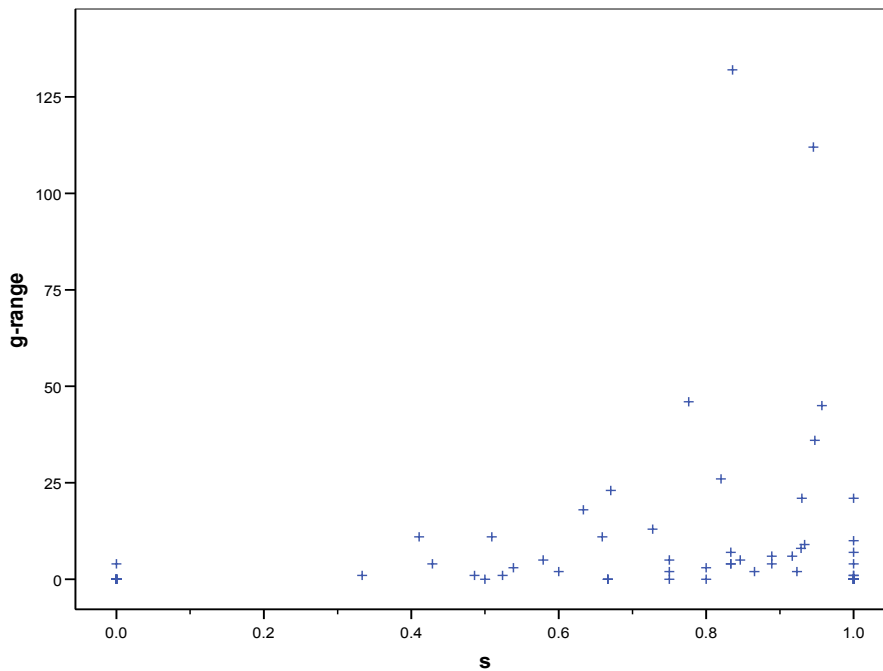
Figure K16. Scatter plot of *g*-ranges against *s*

176

Figure K16 showed no association between *g*-ranges and *s*.

Figure K17. Scatter plot of *g*-ranges against *v*

Figure K17 confirms the lack of any association between *g*-range and *v*.

Figure K18. Scatter plot of h^+ against Λ

177

Correlation of citation metrics of researchers with the Research Assessment Exercise (2001) scores of their employing departments

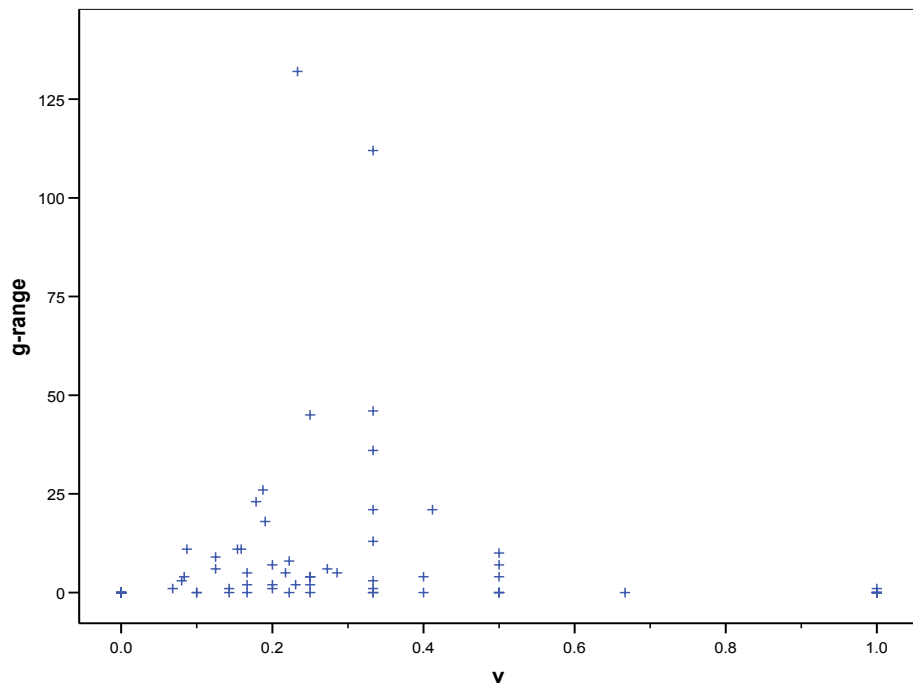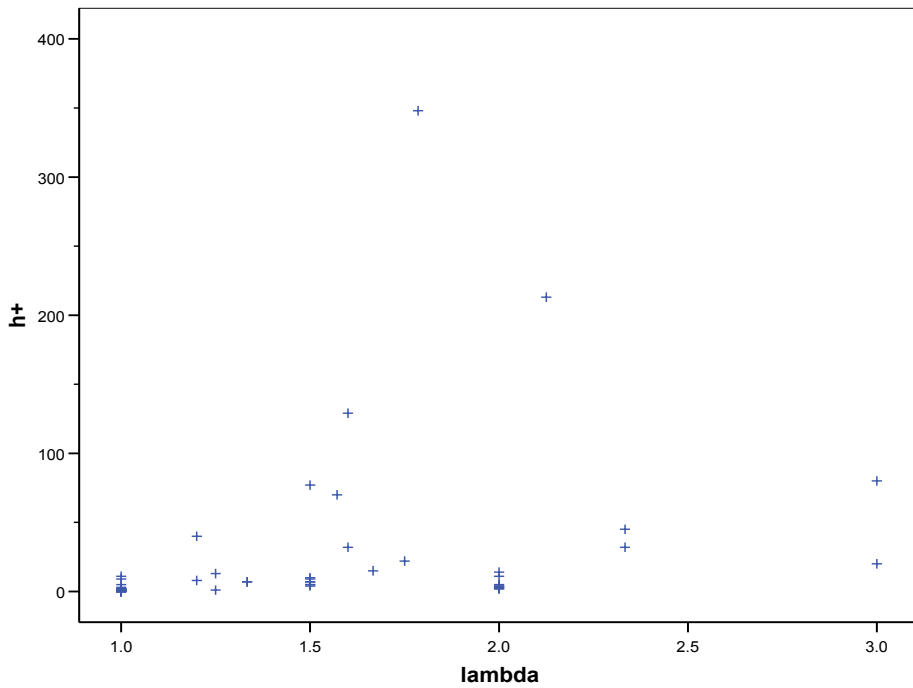|			h	h-range	v	lambda	c_tot	p_tot	Citations_ per_work	RAE_Rank
Spearman's rho	h	Correlation Coefficient	1.000	.771**	.024	.090	.881**	.910**	.734**	.211
		Sig. (2-tailed)	.	.001	.932	.749	.000	.000	.002	.451
		N	15	15	15	15	15	15	15	15
	h-range	Correlation Coefficient	.771**	1.000	.104	.250	.860**	.591*	.689**	.114
		Sig. (2-tailed)	.001	.	.713	.368	.000	.020	.005	.686
		N	15	15	15	15	15	15	15	15
	v	Correlation Coefficient	.024	.104	1.000	-.115	.184	.054	.364	-.027
		Sig. (2-tailed)	.932	.713	.	.682	.511	.849	.182	.923
		N	15	15	15	15	15	15	15	15
	lambda	Correlation Coefficient	.090	.250	-.115	1.000	.204	-.013	.354	-.116
		Sig. (2-tailed)	.749	.368	.682	.	.466	.962	.195	.681
		N	15	15	15	15	15	15	15	15
	c_tot	Correlation Coefficient	.881**	.860**	.184	.204	1.000	.744**	.867**	.246
		Sig. (2-tailed)	.000	.000	.511	.466	.	.001	.000	.376
		N	15	15	15	15	15	15	15	15
	p_tot	Correlation Coefficient	.910**	.591*	.054	-.013	.744**	1.000	.606*	-.052
		Sig. (2-tailed)	.000	.020	.849	.962	.001	.	.017	.854
		N	15	15	15	15	15	15	15	15
	Citations_per_work	Correlation Coefficient	.734**	.689**	.364	.354	.867**	.606*	1.000	.073
		Sig. (2-tailed)	.002	.005	.182	.195	.000	.017	.	.796
		N	15	15	15	15	15	15	15	15
	RAE_Rank	Correlation Coefficient	.211	.114	-.027	-.116	.246	-.052	.073	1.000
		Sig. (2-tailed)	.451	.686	.923	.681	.376	.854	.796	.
		N	15	15	15	15	15	15	15	15

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Appendix M

Please consult CD-ROM (attached inside back cover).

Appendix N

Citations bear a strong resemblance to other indicators of approval in situations where an item or individual can only signal approval once at a time. Citation analysis could be profitably adapted to improve access to archived records in institutional repositories and e-print archives (Harnad *et al.* 2004, p. 314). There are many situations where citation analysis techniques might be adapted to analyse distributions of other variables, for example votes within elections, or borrowing, document supply and inter-library loan statistics within designated collections within a library, where metrics might illuminate different facets of borrowing and use statistics to compare the relative usage of different collections.

Both journal citation counts, use counts and book borrowing counts across a variety of library collections appear to obey the Bradford-Zipf Law across a (Wallace 1987, p. 43). Since libraries justify their expenditure in terms of the level of use of their resources (Nicholas *et al.* 2005, p. 1441), quantitative measures of usage and obsolescence may usefully inform collection management decisions, particularly when choosing which titles to select or deselect from expensive journal bundle packages, and to identify the minimum acceptable access period for a moving wall subscription would be, based upon clients' citation half-life data for each title in the bundle. Citation, use and borrowing metrics could be particularly important for the assessment and ranking of content in open access institutional repositories (OAIRs) in the absence of traditional journal brands if libraries attempt to move to an open-access publishing model (Harnad *et al.* 2004, p. 314).

Subscriptions to less heavily used and cited publications, which Swigger & Wilkes (1991, p. 46) found to be the same titles, although other studies were less convincing (Wallace 1987, p. 46) should arguably be continued in favour of more heavily used titles. Historically, stock evaluation has been qualitative in nature and both library staff and academics chose to deselect journal titles

effectively at random with respect to the number of times academics in the institution cited them (Swigger & Wilkes 1991, p. 46). Arguably, therefore, collection management has not been optimal and both scholars and their advisors have been making erroneous judgements. In the current climate emphasising evidence-based practice, quantitative methods are now necessary and a reliable formula should be researched and promoted to libraries.

The following is a tentative suggestion of how the citation metrics examined earlier in this dissertation might be adapted to assist collection management in a public library.

For library collections, v would indicate the proportion of the collection that was most heavily borrowed. Although somewhat arbitrary, h and v would provide comparable standards that would show not only how much a collection was used but how much the most used books were used, how many books were heavily used and what proportion of the collection this represented.

In addition to citation and total re-shelving counts for each title, new citation metrics may be profitably employed to explore different aspects of collection use. For example, identifying the h most borrowed titles in a collection might help libraries to plan what will require replacement and to suggest which types of book would be required.

In addition to h and v, the following indices, B_n, v_B and r are proposed for library collection management.

B_n = number of books of a certain category of stock, perhaps a subject collection, the volumes of a particular subscribed e-serial title or the titles of an e-serial that are obtained as part of a bundle, or a type of item that has been borrowed (or used) at least B times during the last n years. A similar metric, B_w, is also proposed for shorter time-frames, where B_w = number of books in the specified collection borrowed (or used) at least B times in the last w weeks.

This is a direct application of the *h*-index to use/borrowing statistics. *h* would be less vulnerable to people randomly taking books off the shelves and not reading them than standard "total number of times taken off the shelf" counts. It could however be extended to usage statistics of journals – which years/volumes/issues are used most often, or if just one article is worth obtaining copyright rights for and the rest of the subscription may be stopped.

v_B (*h*/total number of books in the collection) would then be a measure of the relative success of that collection: if the total borrowing (usage) count and/or h_B are high for the collection but v_B is low, this suggests that only a small proportion of the collection is being used but that those books are being borrowed heavily.

Existing measures, such as mean borrowing counts are not easily comparable unless expressed as a proportion of the collection size, *i.e.*:

$$r = \frac{l_c}{n_c}$$

where r = relative borrowing rate of collection

l_c = proportion of all loans that came from collection

n_c = proportion of items in library that are part of this collection

Making the simplifying assumption that all books occupy the same volume of space, r is the relative borrowing rate per book per unit volume allocated to that collection.

The average size of books, expressed as the mean outward facing area of twenty randomly selected books from the collection arranged as on the shelves, may be used as a coefficient to improve the estimated shelf space. The mean cost may be factored in as well. The average size may be calculated by randomly selecting twenty books from the collection, measuring the vertical height of the tallest of them, and multiplying this measurement by the total width of the spines of the

same twenty books arranged side-by-side, as if on display, and dividing this product by twenty.
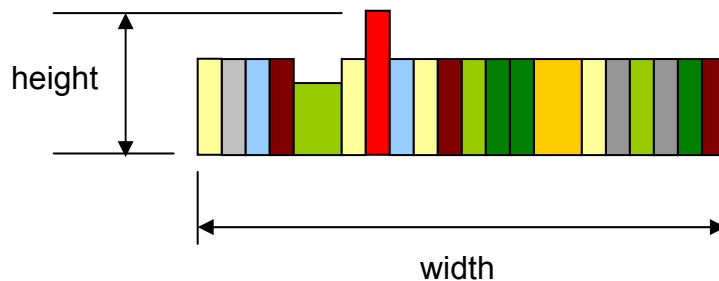
The estimated volume of the collection (calculated by multiplying the average volume of one book from the collection, calculated above, by the number of books in the collection) could then be divided by the volume of space available in the library. Since shelf depths are constant, a suitable index for volume is the sum of shelf-lengths multiplied by the shelf-height. Multiplying this proportion of library space devoted to a collection by r would then give the instantaneous number of borrowings.m^{-2}. Dividing this by the total bought-as-new value of book stock (incorporating inflation) would give the instantaneous number of borrowings.m^{-2}.£$^{-1}$ devoted to stock. This is the best standardised general measurement because it provides an accurate estimate of the value obtained per unit of library space and book fund invested in a collection.

These metrics are rapidly calculated and unlike total borrowing counts or percentage of total borrowing counts, they are comparable and relate to the relative return per unit invested in a collection. Trends in these metrics over time may be compared.

These metrics complement traditional metrics of total borrowing counts and graphs of borrowing counts per item but provide comparable descriptive statistics to help guide collection management.

Appendix O

As works published by an individual become more heavily cited, more individuals are likely to encounter the first author's work through reading the much larger number of works that cite it. Some of these will cite the original work, so provided the bulk of researchers do not cite the secondary source instead of the original, the number of citations to the original work will increase, creating yet more signposts, probably across an increasingly large range of research, making the original publication much more visible by presenting researchers with many signposts to the original. The probability of others then encountering a citation to the original work and citing it therefore increases. Visibility and citedness are therefore tautologically related: increasing citation increases visibility and increasing visibility increases citedness, provided secondary citation is minimal and activity in relevant subfields is constant or increases. This tautology is a philosophical axiom that cannot be tested through citation analysis.

Miscitations may be identified by studying publications by different researchers and counting the number of miscitations to individual works (where the work in question gains citations) and the number of miscitations that should have gone to the work that went to other publications (depriving the work of citations). The total number of citations to a work incorrectly awarded to other works could be subtracted from the number of citations awarded to the work.

This could be repeated for a large number of works for a large number of researchers. A one-sample, two-tailed, Wilcoxon Signed-Rank test would determine whether the median value was significantly different from zero. A significant difference from zero for individual works, or for individual researchers (where the total miscitation differences are summed for all publications), would indicate that the bias was systematic. The distribution of miscitation differences should also be examined to ensure that individuals or publications with distinct characteristics are not selectively affected more or less than average by miscitations, and that systematic effects have not cancelled one

another out overall. Miscitations have been found to follow Bradford-Zipf Law following trends in researchers not adequately reading publications (Simkin & Roychowdhury [n.d.]), so if categories of different types of publication appear to attract a greater proportion of positive or negative miscitations than might be expected by chance, more data should be gathered, classified and this new data tested to identify the trend.

The proportion of secondary citations to citations to the original work could be measured for different subfields. For citation analysis to be valid, this ratio should be low and constant over time.

The most cited authors in different subfields should be identified and the proportion of citations by individuals to authors of different citedness in the subfield in which the publication was published. Rather than assessing research relative to arbitrary categories, as White (2004) did, this would provide relative measures for different subfields. These could then be normalised (scaled) to produce results that could be compared between subfields.

The number of citations to singly and co-authored works (manually excluding obvious secondary citations) could be counted and compared using a χ^2 goodness-of-fit test, with a null hypothesis that singly and collaboratively authored papers would be in a 1:1 ratio. Both sets of publications could then be divided by the number of fields (or even subfields) to which they were relevant, to control for the effect of visibility to multiple fields, and tested again.

A significant result to only the first test would suggest that one form of authorship acts systematically but through a mechanism other than breadth of relevance. A significant result only to the second test would suggest that breadth of relevance systematically determines relevance regardless of authorship, and a significant result to both tests would suggest that the number of authors and number of fields both exert systematic and interacting effects. If neither test is significant, it may be concluded that, given the lack of conclusive evidence that single or co-authorship has a systematic effect increases citedness, that both factors act either randomly or are unimportant.

Appendix P

If researchers within a subfield cite one another more frequently than they cite those outside the subfields that they research, subfields may be defined by the number of citations between researchers. The following is an outline of a crude analytical method that would help to tentatively identify subfields and individuals that might contribute to them with limited or no prior knowledge of subfield diversity. The involvement of a computer programme capable of interrogating a citation database would be essential.

1. Count the number of citations between researcher A and each of the researchers that they cite.

2. Repeat for other individuals – there are two ways of achieving this:

    (i) Count the number of citations between each person the first individual (selected at random) cited, as in step 1.

    (ii) Repeat with other randomly selected individuals

3. Where more than n citations in both directions exist between individuals, they may be considered part of a group researching similar materials. From here on such groups will be described as "pseudo-subfields".

4. Once all pseudo-subfields have been identified, by examination of the research interests and authorised publication lists of individuals within each group, they may be grouped into true subfields. It is likely that some individuals will be left out of relevant pseudo-subfields using this approach and some researchers included inappropriately. These errors may be corrected at the qualitative refinement stage.

5. n will be pseudo-subfield specific and will have to be decided upon arbitrarily based upon a study of the work of individuals certain to be in the field (number of citation linkages >> proposed n) and those on the periphery.

6. n should begin as large as possible to define the most productive pseudo-subfields and then be reduced in size, stepwise, until new distinct clusters

emerge that become obscured by noise if *n* is reduced further. At each stage, established pseudo-subfields should be demarcated and all contributing linkages ascribed to these subfields hidden. Linkages between individuals representing other subfields would then become more obvious allowing these to be identified.

7.   Significant numbers of citations in both directions between members of one identified subfield research community to other clusters of mutually citing researchers would indicate membership. All others linked to that individual might also be members of the second subgroup because some of the citations describing the first subgroup may have been describing the second subgroup as well if members were part of several subgroups – at this point it would be necessary to consult the researchers declared research interests or to contact them.

Any level of reciprocal citation indicates possible subgroup membership. Heavy one-way citation might indicate a younger researcher citing a prestigious researcher but small numbers of one-way citations might indicate citation of related research in a different subfield, *i.e.* that the recipient of the citations may be influential but not a member of the citing researchers' subgroup, or they may be leading an emerging research front.

It would be very difficult to identify different pseudo-subfields which comprised the same researchers because it would be functionally impossible to distinguish between citations relating to the different pseudo-subfields without exhaustive research.

The analysis should ideally be repeated (ideally by different researchers) several times using different starting individuals. The same clusters should become apparent. Automation of the process would be useful, as would development of an algorithm predicting the probability that an individual belongs to a particular pseudo-subfield.

Different crude clusters could be suggested and a computer could then calculate which clusters have the greatest number of reciprocal citations. It might be possible to statistically determine the solution that maximises the average number of citations in both directions between each publication and their membership clusters.

Comparison with co-citation analysis and bibliographic coupling

Co-citation analysis clusters highly cited and co-cited documents with one another to form aggregates of densely networked publications that are consistently cited together by other publications and represent the "intellectual base" of a field or subfield, to which less highly cited (possibly more recent) publications may cite, which represent the current research front (Jarneving 2005, p. 248). Over time, some of these publications would become heavily cited and co-cited by other contemporary documents, and thus come to form part of the expanding intellectual base of the subfield.

Bibliographic coupling clusters publications that cite the same works, rather than cited publications. Bibliographic coupling establishes a research front that does not move because the citations that define it remain forever, but which expands as more publications cite the same base literature. (Jarneving 2005, p. 246)

Jarneving (2005, p. 254) found that, as expected after Seglen (1992), very few heavily cited publications contribute to these citation fronts. The two methods also produced very different research front maps from one another (Jarneving 2005, p. 254). The understanding of the application of these methods appears to be in its infancy.

The proposed method is similar to co-citation analysis but instead of mapping research fronts and discrete clusters of established theory, it maps the flow of information between pseudo-subfields and the intellectual and social associations between researchers. By focussing on individuals rather than publications, this method would indicate the strength of associations both between subfields and

between individuals, allowing tentative mapping of social networks and connections between teams that might be used to inform future studies of citation behaviour. It would be certain to identify cronyism, reciprocal and group citation behaviour.

Individuals leading research fronts would probably be easier to define using a method similar to bibliographic coupling. Instead of counting numbers of citations between individuals, the number of citations to and from all the other researchers could be compared for each individual. Individuals who cite and are cited by the same individuals are more likely to be involved in the same subfields.

By studying the research interests of individuals involved in different fields, both methods would show which subfields are studied by the same individuals, and which are therefore presumably similar or related in some way. This might suggest opportunities for using methods and concepts from one subfield in another.