# Benchmarking citation measures among the Australian education professoriate

Peter R Albion

University of Southern Queensland

Peter Albion is an Associate Professor in the Faculty of Education at the University of Southern Queensland, Toowoomba. He teaches and researches in areas related to information and communication technology in education, including online education.

# Benchmarking citation measures among the Australian education professoriate

**Abstract**:

Individual researchers and the organisations for which they work are interested in comparative measures of research performance for a variety of purposes. Such comparisons are facilitated by quantifiable measures that are easily obtained and offer convenience and a sense of objectivity. One popular measure is the Journal Impact Factor based on citation rates but it is a measure intended for journals rather than individuals. Moreover, educational research publications are not well represented in the databases most widely used for calculation of citation measures leading to doubts about the usefulness of such measures in education. Newer measures and data sources offer alternatives that provide wider representation of education research. However, research has shown that citation rates vary according to discipline and valid comparisons depend upon the availability of discipline specific benchmarks. This study sought to provide such benchmarks for Australian educational researchers based on analysis of citation measures obtained for the Australian education professoriate.

## Introduction

Comparisons of research outputs are made for various reasons. Governments seeking a basis for allocation of limited research funds have developed schemes such as the *Research Assessment Exercise* (RAE) in the United Kingdom (Moed, 2008), the *Performance Based Research Fund* (PBRF) in New Zealand (Hodder & Hodder, 2010; Smith, 2008), and the Australian Government's *Excellence in Research for Australia* (ERA) scheme (Australian Research Council, 2009). Universities and other research organisations also have an interest in selecting and supporting the best available researchers.

Research is a complex activity and assessment of its quality needs to consider that breadth and not be dependent upon a single measure. However, the schemes referred to above use bibliometric statistics based on citation data as an indicator of quality. Justification for using citation rates to indicate research quality is based on publication being a "key component of the social system of science" (Herther, 2009, p. 363) and the role of citation as an indicator of use and, hence, the impact of the research (Bornmann, Mutz, Neuhaus, & Daniel, 2008).

Research on citing behaviour has found that decisions to cite a work may be affected by a variety of reasons. Bornmann and Daniel (2008) reviewed approximately 30 studies of citing behaviour published over a period of about 40 years. They identified eight types of citation and concluded that caution was required by those interpreting citation and that citation measures should be used as just one indicator of quality of research.

Garfield proposed the ISI (Institute for Scientific Information) Impact Factor (IF) in 1979 as an indicator of quality of journals (Moed, 2005). However, "the journal statistics [Garfield] derived were soon isolated from the study context and published by ISI in rankings of journals by impact factor, probably the bibliometric construct most widely used in the scholarly and publishing community" (Moed, 2005, p. 13). This single measure has been seen as an indicator of the quality of journals and, by extension, of the output of those who publish in them.

Moed (2005) noted that assessing research output of researchers or groups using the IF is subject to risks associated with possible errors in data collection, with the use of simple statistics rather than more sophisticated ones that reduce the effect of biases, and with the lack of valid reference values for comparison. The purpose of this paper is to address the latter limitation for Australian educational researchers.

## Journal impact factor

According to Moed,

> The impact factor of a journal J in year T is defined as follows:
>
> | |
> |---|
> | The number of citations received in year T by all documents published in J in the years T-1 and T-2 |
> | ÷ |
> | The number of citable documents published in J in the years T-1 and T-2 |
>
> (Moed, 2005, p. 92)

The calculation of impact factor as a ratio accounts for differences in the numbers of articles published in different journals but does not account for discipline-based differences in citing behaviour or for differences in the typical time between publication of an article and its inclusion in reference lists. Both of these differ substantially between disciplines, making comparison of IF across disciplines unreliable (Moed, 2005).

An Australian study (Jarwal, Brion, & King, 2009) found wide variations by discipline, with journal IF ranging from 3.37 to 30.3 for biological sciences and 0.68 to 51.30 for clinical sciences and clinical physiology. The IF range for economics, commerce and management was 0.35 to 1.43.

Education journals rank in the lower third of social sciences journals with a mean IF of 0.70 compared to 1.77 for psychiatry, 1.50 for clinical psychology, 0.91 for law and 0.75 for sociology (Goodyear, et al., 2009). For a core set of 11 education journals nominated by expert panels, Goodyear et al. reported IF values from 0.14 to 2.60, with minimum and maximum considerably lower than reported for other disciplines (Jarwal, et al., 2009).

A particular difficulty for using IF in education is the limited coverage of relevant education journals in the ISI databases. Database "coverage of the *journal* literature is in most main fields *excellent* or *very good*, except for those parts of social sciences as sociology, education, political sciences and anthropology, and particularly for humanities & arts" (Moed, 2005, p. 135).

Database coverage varies by country as well as by discipline (Moed, 2005). In an Australian Government sponsored bibliometric analysis of the international contribution of Australian educational research (Phelan, Anderson, & Bourke, 2000) Australian researchers performed comparatively well but 75% of their publication output was in local journals not listed in ISI. Recognition that the traditional measures of research quality using citation frequency and journal impact factors reflect inadequate coverage of European publications, especially in

languages other than English, is a key driver of the *European Educational Research Quality Indicators* (EERQI) project funded by the European Commission (http://www.eerqi.eu/).

The IF is an indicator of the impact of a journal rather than of an individual paper or researcher. Analysis of citations during 2004 of papers published in Nature during 2002 and 2003 found that 89% of the IF (32.2) was generated by 25% of the papers (Campbell, 2008). Within a journal with a high IF there can be considerable variability in the citation rates for individual papers and, by extension, researchers. It is not reasonable to extend judgment about the quality of a journal based on IF to individual papers or researchers who publish in the journal.

## Citation measures for researchers

The IF is not appropriate for use as an indicator of quality of output from a researcher. If an index for researchers based on citation rates is to be used then it should be one designed for that purpose.

Of alternative indicators proposed for individual researchers, probably the best known is the h-index as defined by Hirsch: "A scientist has index $h$ if $h$ of his or her $N_p$ papers have at least $h$ citations each and the other ($N_p$ - $h$) papers have ≤$h$ citations each" (Hirsch, 2005, p. 16569). The $h$ index can be easily derived by rank ordering papers according to the number of citations.

Since the $h$ index was first proposed (Hirsch, 2005), other researchers have proposed variations to overcome perceived disadvantages. Bornmann and Daniel (2009) mention several, including the $g$ index which better represents highly cited papers, the $m$ quotient which adjusts for the length of time since the first published paper and the $h_i$ index which accounts for co-authorship.

Bornmann and Daniel (2007) concluded that the $h$ index provides a robust indication of cumulative productivity of a researcher and is insensitive to both lowly cited papers and a small number of highly cited ones. However, they noted that its cumulative nature means that it favours enduring performance and it may be unsuitable for comparing researchers at very different stages in their careers.

Although the $h$ index appears better suited than IF for individual researchers, there are disadvantages (Panaretos & Malesios, 2009). It is bounded by the total number of publications and so disadvantages new researchers, however significant their work. It is affected by self citations, has slightly less predictive accuracy than mean citations per paper, disadvantages small but highly cited outputs, suffers from confusion of similar names, is affected by limitations in the databases used, and is prone to the problems of over-simplification through using a

single measure. “Overall, as a general guideline for assessing the citation impact of a researcher, [they] suggest a combined use of the h-index with other h-type indices for more representative results” (Panaretos & Malesios, 2009, p. 666).

One of the “other h-type indices” that appears to have gained comparatively widespread recognition is the *g* index:

> A set of papers has a *g*-index *g* if *g* is the highest rank such that the top g papers have, together, at least $g^2$ citations. This also means that the top $g + 1$ papers have less than $(g + 1)^2$ papers (Egghe, 2006, p. 132)

This formulation accords more weight to the citations of the most highly cited papers in excess of the number needed to contribute to the *h* index. In this way it “resembles more the overall feeling of ‘visibility’ or ‘life time achievement’” (Egghe, 2006, pp. 142-143) of a researcher.

Using data from the 2008 UK Research Assessment Exercise, Norris and Oppenheim (2010) examined the correlation between the *h* and *g* index values and rankings by peer assessment and between the RAE rankings and the collective *h* and *g* index of submitting departments. They found that the correlations varied by discipline, being strong for pharmacy, less strong but still reasonable for library and information science, and inconsistent for anthropology. The data source used for citations was WoS and the more limited representation of anthropology in that database (Moed, 2005) probably accounts for the result.

## Sources of data for citation analysis

Because the ISI database on which the IF was originally based has, over time, become part of the Thomson Reuters Web of Knowledge, which includes the (WoS), the references in the literature variously refer to ISI and WoS as sources of the data used to derive the IF. Scopus, published by Elsevier, is now an established alternative to WoS for accessing citation records and is the officially selected source to be used in the ERA (Australian Research Council, 2009). Both WoS and Scopus are paid services and each restricts its analysis to the journals indexed in the database. Consequently there are differences in the measures obtained using the two systems because the sources they index overlap but do not coincide (Meho & Yang, 2007).

Compared to WoS and Scopus, Google Scholar offers advantages in cost (free) and breadth of coverage (the entire Internet) at the expense of the inclusion of fringe material. However, the ready availability of free tools such as *Publish or Perish* (Harzing, 2009) and *Scholarometer* (http://scholarometer.indiana.edu/) that allow

direct calculation of *h* and *g* index values makes Google Scholar an attractive alternative (Harzing & van der Wal, 2008).

Meho and Yang (2007) compared citation counts for scholars in library and information science using WoS, Scopus and Google Scholar. They obtained a high correlation (0.97) between the citations found in Google Scholar and those found in the union of WoS and Scopus. This study used raw citation counts rather than the *h* index, which is less dependent on locating all citations, and did not use automated systems such as *PoP* (Harzing, 2009) or *Scholarometer*.

Bar-Ilan (2008) compared the *h* index values derived using WoS, Scopus, and Google Scholar for highly cited Israeli science researchers and reported substantial variations by data source depending on the discipline. Some differences were explained by differences in self-archiving of documents that are then available to Google Scholar but are not indexed by WoS or Scopus.

In another study (Vaughan & Shaw, 2008) that compared WoS, Google and Google Scholar, the Google Scholar citations fell between those recorded for the tightly controlled WoS and the uncontrolled Google. Correlations of Google Scholar citations with WoS ranged between 0.43 and 0.75 depending on the type of citation or publication and 92% of the citations returned by Google Scholar showed intellectual impact. The researchers concluded that Google Scholar has potential to be a useful tool in research evaluation.

A study comparing WoS with Google Scholar for citations in the area of management and international business found that Google Scholar resulted in more comprehensive citation coverage and benefited academics published in sources not well covered by ISI/WoS, such as books, conference papers and non-US journals (Harzing & van der Wal, 2008). Values obtained for *h* and *g* index and citations per paper using Google Scholar correlated strongly with IF and offered advantages including availability without cost.

## Variability of citation measures across disciplines

Citation practices vary across disciplines (Moed, 2005) and studies comparing measures such as IF across disciplines have found considerable variation (Jarwal, et al., 2009). Studies of alternative measures such as the *h* index have reported similar variation. As a consequence, researchers can calculate their own *h* and *g* index scores but, in the absence of values for their peers, are unable to obtain useful indications of relative standing.

A study benchmarking Italian science researchers on *h* and *g* index scores over a 5-year window from 2001 to 2005 examined data for 27000 researchers in 165 discipline areas (in 9 broad groups) across 79 universities

(Abramo, D'Angelo, & Viel, 2010). Median values for *h* ranged from 2 to 6 and for *g* from 3 to 8. Mean values ranged from 2.31 to 6.24 for *h* and 3.38 to 9.18 for *g* and maximum values were 36 for *h* and 58 for *g*.

A study of Australian information systems researchers (Clarke, 2009) compared Thomson/ISI citation counts and the *h* index calculated using *Publish or Perish* (Harzing, 2009) and concluded that, at the end of 2007, appropriate benchmarks might be a *h* index of 25 (with a total of 750 citations) for an *outstanding* Australian IS researcher and a *h* index of 12 or 15 (with a total of 500 citations) for a *successful* Australian IS researcher.

Top performing researchers published in four premier marketing journals had *h* index scores ranging from 3 to 17 with median values of 9 to 11 (Saad, 2010). By comparison, top performers in other business related areas had *h* index scores ranging from 9 to 24. The implication is that even within related areas there is considerable variation according to discipline.

In another study of the stability of the *h* index, scores for 5614 computer scientists were found to have an average *h* index of 2.19 and a median of 1 when derived from ISI data and an average of 3.54 and a median of 2 when derived from Google Scholar (Henzinger, Suñol, & Weber, 2010). By comparison, 1375 physicists were found to have an average *h* index of 7.15 and a median of 3 using ISI and an average of 6.70 and a median of 4 using Google Scholar.

The studies cited above (Abramo, et al., 2010; Henzinger, et al., 2010; Saad, 2010) demonstrate the variability of the *h* index according to discipline and support the assertion of Bornmann and Daniel (2009) that, if the *h* index is to be used to evaluate research performance, it should be used for researchers of similar career length and in the same field of study. Thus the utility of the *h* index as a gauge of research performance depends upon the availability of relevant benchmarks for comparison.

## Benchmark index scores for Australian educational researchers

Implicit in the discussion above is the conclusion that, if citation measures are to be used for evaluation of research performance, they should be used in ways that ensure comparison of like researchers. The context of Australian educational research is sufficiently different, even from other parts of the English speaking world, in respect of career trajectory and resourcing to justify treating it, rather than a more international selection as the basis of comparison for Australian educational researchers.

Constructing a list of these researchers would be no easy task because they are typically not gathered into single organisational section of a university. In the 2010 ERA exercise, which assessed research according to Field of

Research (FoR) codes, a considerable number of publications coded as educational research (FoR 13) came from researchers in other disciplines who had published research related to teaching in their core discipline or from researchers in sections of universities supporting teaching functions rather than faculties or schools of education. Equally, some researchers in faculties and schools of education had published work linked to other FoRs. The broad category of educational researchers is permeable and its membership is constantly changing.

Identifying and calculating *h* index scores for all Australian educational researchers would be a challenging task and the result would not meet the second criterion of similar career length suggested by Bornmann and Daniel (2007). Hence it is desirable to identify a smaller set of educational researchers expected to have broadly comparable career lengths. The Australian education professoriate, university academics holding positions as professors and associate professors in Faculties of Education (or equivalent), presents as an identifiable group appointed to positions that typically include an expectation of research performance. This group will typically represent the most experienced and successful educational researchers and benchmarks derived from their performance should be indicative of strong research performance. The effect of continuing increase in *h* index even beyond active publishing can be eliminated by excluding emeritus, adjunct and other forms of appointment likely to be occupied by researchers at or beyond the typical limits of a research career. This approach may exclude some researchers who would have been significant contributors of educational research for the ERA and subsequent research might seek to extend the range of coverage.

Neither the traditional source of citation data (Moed, 2005), ISI WoS, nor that selected for the ERA (Australian Research Council, 2009), Scopus, provides a strong representation of the citation data for educational research (Bates, 2003; Levine-Clark & Gil, 2009; Moed, 2005; Phelan, et al., 2000). Hence Google Scholar is likely to provide a more suitable source with broad coverage of the field. The risks associated with the inclusion of less authoritative material should be balanced by the more comprehensive coverage, especially of Australian publications, which are not well covered in the conventional sources (Phelan, et al., 2000).

Although the *h* index is the simplest to obtain and provides a useful indication of the broad impact of a researcher's work (Hirsch, 2005), it is known to under-value publications that accumulate more citations than are required for them to be included in the group that contribute to the *h* index. The *g* index (Egghe, 2006) compensates for this by attributing more weight to highly cited publications and would be a useful additional benchmark.

The overall question to be answered by this attempt to develop citation index benchmarks for Australian educational researchers is:

What *h* and *g* index values represent strong research performance by an Australian educational researcher?

This overall question can be considered in relation to a set of subsidiary questions:

1. How are *h* and *g* index values distributed within the Australian education professoriate?
2. How do the distributions of *h* and *g* index values among the Australian education professoriate differ for professors as compared to associate professors?
3. How do the distributions of *h* and *g* index values among the Australian education professoriate differ for universities that belong to identifiable groups?
4. What *h* and *g* index values might be offered as indicative benchmarks for members of the Australian education professoriate?

## Methodology

Universities Australia (http://www.universitiesaustralia.edu.au/) has a web page listing member universities with links to their websites. This page was used as the starting point for data collection during January and February 2010. The website of each university was visited and searched for indications that the university offered studies in education. Organisational units offering studies in education were found for 35 of the 39 universities listed by Universities Australia. For each such unit a list of currently active professors and associate professors was compiled in a spreadsheet using the data available from the website. Staff indicated as holding emeritus, adjunct, honorary or similar positions were excluded from the list.

The Universities Australia website lists three identified groups of universities, Group of Eight (Go8, 8 members), Australian Technology Network (ATN, 5 members), and Innovative Research Universities (IRU, 7 members). Membership of these groups was recorded for subsequent use in analysis of whether universities in the groups rated differently on the measures being investigated.

Harzing's *Publish or Perish* (PoP) software (Harzing, 2009) was used to obtain citation records and index values from Google Scholar. For simplicity, data were collected for all identified researchers at one university before moving on to another university. This approach facilitated using information on the university website such as lists of publications for checking the publication data returned by Google Scholar.

PoP was set up with the "Social Sciences, Arts, Humanities" category selected and the other categories deselected. For each researcher data collection began with entry of the first and last name of the researcher in the search field of PoP. The software returned a list of papers attributed to the author, ranked by number of recorded citations and displayed in a table with author(s), title, year of publication, publication and publisher. Some entries had one or more of year, publication, and publisher blank. Each entry in the list had a checkbox that could be toggled to exclude (or include) that entry. Above the list, a table displayed a selection of statistics including number of publications, number of citations, *h* and *g* index values.

Where the number of entries returned seemed abnormally low or high, a variation of name, such as substituting a full first name for a diminutive or using an initial rather than the full name was tried. Where the university website provided a list of publications they could be checked for the correct variant of the name, or an alternative in the case of name change, to try. In general this process was repeated once or twice to maximise the number of entries in the pool. Once a sufficiently large pool of publications was obtained it was checked to remove irrelevant entries. The list was sorted by date and very old or undated entries were removed by toggling the checkbox. The list was then sorted again by citation count from highest to lowest and each entry was scanned to check that the researcher was in the list of authors and that the title of the paper indicated a field of research consistent with the work of the researcher listed on the university website. Entries that were judged not to belong to the researcher were unchecked. Working down the list of entries reduced the initial values of *h* and *g* as entries were unchecked and the process was halted once the index values ceased to change.

The process of obtaining index values was reasonably straightforward for most researchers. Difficulties were encountered with researchers who had changed name and with some others for whom the searches did not return articles that were listed on the relevant university website. Where the data appeared to be unreliable the researcher was not included in further analysis.

## Results

A total professoriate of 411 members comprising 194 professors and 217 associate professors was identified across the 35 universities for which indications of studies in education were found. Of these, citation records for five, one professor and four associate professors, appeared to be too few for confident analysis and they were excluded. Data collected in a spreadsheet were transferred to SPSS 18 for analysis.

The distributions of *h* index scores for both professors and associate professors were strongly positively skewed with a small number of high scores resulting in means that were higher than the median (Q2) values. Values

calculated for skewness and kurtosis using SPSS are included in Table 1 which summarises *h* index statistics for professors and associate professors by university groups as well as for the complete data sets for professors and associate professors.

**Table 1: Summary of *h* index statistics by academic rank and university group**

| | N | Min. | Q1 | Q2 | Q3 | Max. | Mean | SD | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| **Professors** | | | | | | | | | | |
| Group of Eight | 54 | 4 | 8 | 11 | 16 | 44 | 12.7 | 6.8 | 2.06 | 7.25 |
| Australian Technology Network | 32 | 3 | 5 | 9 | 13 | 42 | 11.1 | 8.9 | 2.09 | 4.62 |
| Innovative Research Universities | 36 | 2 | 5 | 7 | 13 | 27 | 9.2 | 6.0 | 1.27 | 1.29 |
| Ungrouped | 71 | 2 | 6 | 8 | 12 | 28 | 9.6 | 5.5 | 1.16 | 1.31 |
| All professors | 193 | 2 | 6 | 9 | 13 | 44 | 10.6 | 6.7 | 1.82 | 5.11 |
| **Associate Professors** | | | | | | | | | | |
| Group of Eight | 44 | 2 | 4 | 6 | 10 | 18 | 7.1 | 3.6 | 0.82 | 0.73 |
| Australian Technology Network | 31 | 1 | 4 | 6 | 9 | 17 | 7.0 | 3.4 | 1.03 | 1.37 |
| Innovative Research Universities | 27 | 2 | 4 | 5 | 8 | 10 | 5.5 | 2.4 | 0.22 | -1.14 |
| Ungrouped | 111 | 1 | 4 | 5 | 7 | 21 | 5.7 | 3.0 | 1.43 | 5.21 |
| All associate professors | 213 | 1 | 4 | 6 | 8 | 21 | 6.2 | 3.2 | 1.17 | 2.68 |

Figure 1 shows box plots for the *h* index scores of professors and associate professors arranged by university group and for the complete data sets for professors and associate professors. The box plots have been drawn such that the box represents the second and third quartiles of each distribution, with the line across the box representing the median, and the 'whiskers' extending from minimum to maximum value in each distribution. Checking during data collection confirmed that the high *h* index scores in the long tail were genuine, representing researchers who had unusually high numbers of frequently cited publications.

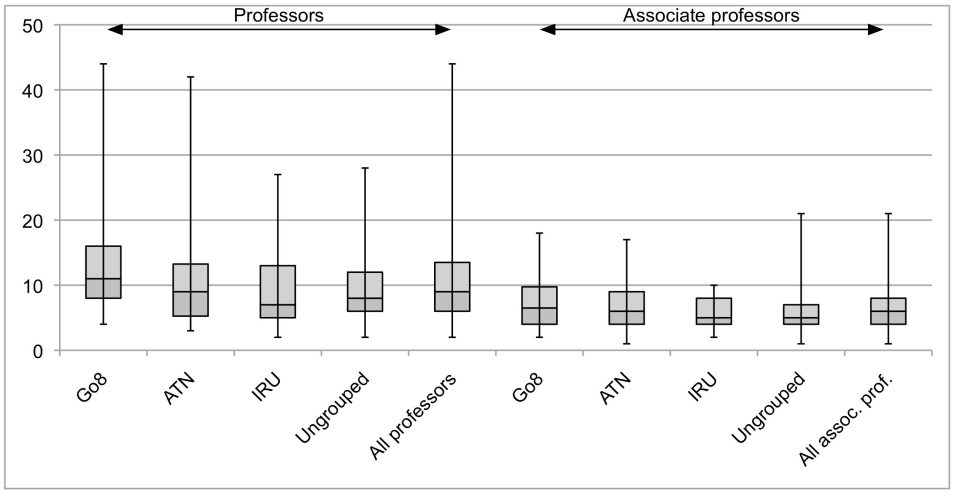


**Figure 1: Box plots of *h* index scores for professors and associate professors by university group**

The *g* index scores produced distributions that were more strongly positively skewed than the corresponding distributions for *h* index scores. Table 2 summarises *g* index scores in a format similar to the summary of *h* index scores in Table 1. Figure 2 presents box plots for the *g* index scores using a format similar to that used for the *h* index scores in Figure 1.

**Table 2: Summary of *g* index statistics by academic rank and university group**

| | N | Min. | Q1 | Q2 | Q3 | Max. | Mean | SD | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| **Professors** | | | | | | | | | | |
| Group of Eight | 54 | 5 | 14 | 23 | 29 | 105 | 24.8 | 15.4 | 2.85 | 13.16 |
| Australian Technology Network | 32 | 5 | 9 | 14 | 23 | 72 | 19.1 | 15.0 | 2.03 | 4.55 |
| Innovative Research Universities | 36 | 4 | 8 | 13 | 21 | 52 | 16.2 | 12.4 | 1.52 | 1.86 |
| Ungrouped | 71 | 2 | 10 | 14 | 21 | 81 | 17.5 | 13.2 | 2.48 | 8.39 |
| All professors | 193 | 2 | 10 | 16 | 25 | 105 | 19.6 | 14.3 | 2.29 | 8.28 |
| **Associate Professors** | | | | | | | | | | |
| Group of Eight | 44 | 2 | 8 | 12 | 16 | 38 | 12.8 | 7.2 | 1.37 | 2.71 |
| Australian Technology Network | 31 | 2 | 6 | 11 | 14 | 34 | 11.7 | 7.0 | 1.55 | 3.14 |
| Innovative Research Universities | 27 | 2 | 5 | 7 | 13 | 16 | 9.0 | 4.7 | .24 | -1.06 |
| Ungrouped | 111 | 1 | 6 | 9 | 13 | 35 | 9.9 | 5.8 | 1.48 | 4.36 |
| All associate professors | 213 | 1 | 6 | 10 | 14 | 38 | 10.6 | 6.3 | 1.46 | 3.58 |

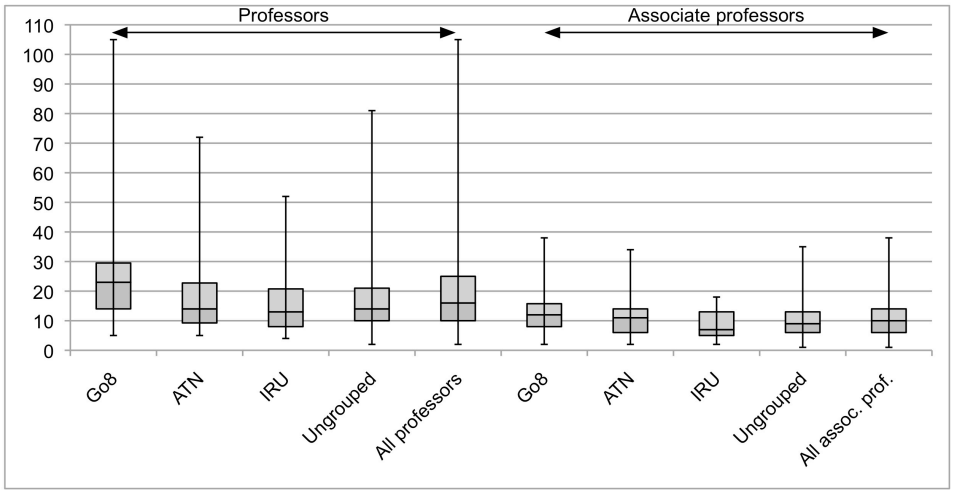


**Figure 2: Box plots of *g* index scores for professors and associate professors by university group**

SPSS 18 was used to calculate separate two-way analyses of variance with *h* and *g* index scores as the dependent variables. Academic level (professor or associate professor) and university group (Group of Eight, Australian Technology Network, Innovative Research Universities, and ungrouped) were the between-subjects variables. For both *h* and *g* index scores there were significant main effects for both academic level and university group. There were no significant interactions between the factors.

The *h* index scores were found to be significantly higher, $F(1, 398) = 59.73$, $p < .001$, for professors ($M = 10.6$, $SD = 6.7$) than for associate professors ($M = 6.2$, $SD = 3.2$). The effect of university group was also significant, $F(3, 398) = 5.16$, $p = .002$, with the means and standard deviations as shown in Table 1. Pairwise comparisons revealed significant differences between Go8 and each of IRU ($p = .002$) and ungrouped ($p = .001$). The other differences between groups were not statistically significant. Figure 3 plots the mean *h* values by academic level and university group.

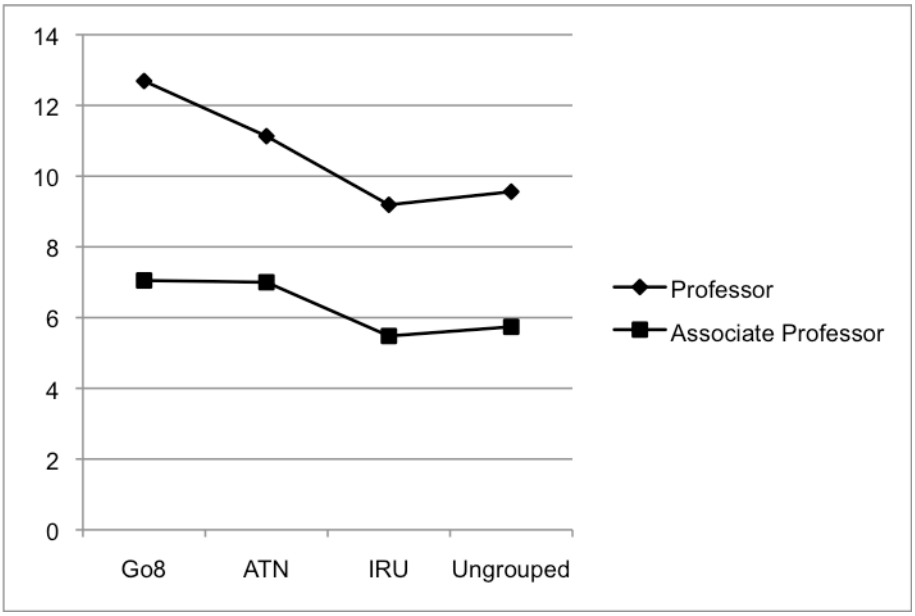


**Figure 3: Mean *h* index values by academic level and university group**

Values for the *g* index were also found to be significantly higher, $F(1, 398) = 53.88$, $p < .001$, for professors ($M = 19.6$, $SD = 14.3$) than for associate professors ($M = 10.6$, $SD = 6.3$) and there were significant differences between university groups, $F(3, 398) = 6.06$, $p < .001$, with the means and standard deviations as shown in Table 2. Pairwise comparisons revealed significant differences between Go8 and each of the other groups (ATN: $p = .050$, IRU: $p < .001$, ungrouped: $p < .001$). Other differences between groups were not significant. Figure 4 plots the mean *g* index values by academic level and university group.

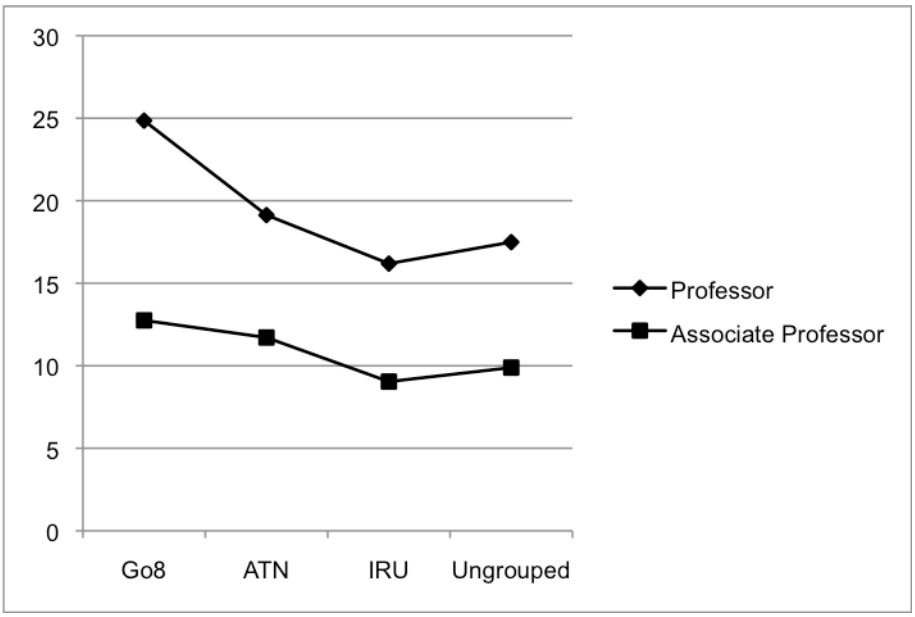


**Figure 4: Mean *g* index values by academic level and university group**

## Discussion

The distributions of *h* and *g* index values within the Australian education professoriate were found to be positively skewed because of the presence of a relatively small number of researchers with larger numbers of highly cited papers. Across the entire data pool, the differences between median and mean for associate professors were minor (6 to 6.2 for *h* and 10 to 10.6 for *g*) but more pronounced for professors (9 to 10.6 for *h* and 16 to 19.6 for *g*). The differences for associate professors are small enough not to raise doubts about use of the mean as a basis for comparison but for professors the differences are large enough to make the median more appropriate than the mean if a representative value of the *h* or *g* index is wanted for purposes of comparison.

There are statistically significant differences between distributions of *h* and *g* index scores retrieved for professors and associate professors. Consistent with their typically longer careers, which naturally produce increases in *h* and related index scores (Hirsch, 2005), and selection or promotion on the basis of research performance, professors have mean *h* and *g* scores that are significantly higher than those of associate professors.

The claims of the Group of Eight universities to comparative excellence in research appear to be supported by this study. The only statistically significant differences by university group were those between the Go8 and IRU and ungrouped universities for *h* index scores and the Go8 and all three other groupings for *g* index scores.

In proposing benchmarks for comparison of researchers on the citation measures considered in this study, it is prudent to consider the median values rather than the means as indicators of typical scores. The most appropriate indicators above and below the median may be the first and third quartile values as used for the lower and upper boundaries of the box plots in Figures 1 and 2. Thus, a typical (median) value is situated within a range encompassing 50% of the relevant population. Table 3 presents three values for each of *h* and *g* for professors and associate professors. The median value is indicated as *typical* and the other quartile values are indicated as *marginal* and *superior*. A professor or associate professor recording an index score within the relevant range can be considered to be performing appropriately on these citation measures. A current appointee, or applicant, recording values at or below the marginal value might be considered to be under-performing on these citation measures, prompting careful assessment of other available indicators for confirmation or contradiction. Similarly, an appointee or applicant recording values at or above the superior value could be considered to be performing beyond expectations on these citation measures.

**Table 3: Indicative benchmarks for the Australian education professoriate**

| | Marginal | Typical | Superior |
|---|---|---|---|
| **Professors** | | | |
| *h* index | 6 | 9 | 13 |
| *g* index | 10 | 16 | 25 |
| **Associate professors** | | | |
| *h* index | 4 | 6 | 8 |
| *g* index | 6 | 10 | 14 |

Finally, if we accept that professors are, by nature of their appointment, generally representative of strong research performance within their field, then *h* and *g* index values equal to or better than the typical values for professors ($h = 9$ and $g = 16$) should represent strong research performance, at least so far as these measures are appropriate.

## Conclusion

Although the conventional sources of citation data, ISI/WoS and Scopus, are limited in their coverage of educational research, it has been possible to use the freely available Google Scholar and the free *Publish or Perish* software to derive alternative measures that are arguably better suited than IF as indicators of the impact of individual researchers. Using these measures, the *h* and *g* index scores for members of the Australian education professoriate have been collected and used to develop benchmarks that might be useful to both individual researchers seeking a comparative assessment of their own performance and organisations assessing research performance for a variety of purposes.

The cautions raised by researchers in the field of bibliometrics should be attended to by anybody seeking to apply these benchmarks. Whatever the importance of citations as an indicator of the impact of research, they are just one indicator. Critical judgments should not be based on a single piece of evidence but should consider a range of available indicators. Although it will always be tempting to supplement publication and citation data with other quantitative indicators such as value of funding attracted and numbers of research students graduated, the EERQI project (http://www.eerqi.eu/) is investigating the use of new technologies to develop content-based indicators which may eventually contribute to a more holistic view of research quality.

The benchmarks proposed in this paper have been developed at a point in time and it is known that *h* and *g* index scores increase because of the accumulation of citations with the passing of time. This would result in a steady increase in the benchmarks from year to year if the composition of the professoriate was constant but it is not. Each year some senior members retire and others are appointed at more junior levels. Even if the numbers retiring and joining are unequal so that the size of the professoriate increases or decreases, it is possible that the replacement of more senior members with typically higher citation counts by more junior members with fewer citations might balance the otherwise inevitable increase in median and mean index scores. Further research in the form of future audits of citations for the professoriate should settle this question and add to the reliability of the proposed benchmarks for assessing relative research performance.

## References


Abramo, G., D'Angelo, C. A., & Viel, F. (2010). A robust benchmark for the h- and g-indexes. *Journal of the American Society for Information Science and Technology, 61*(6), 1275-1280.

Australian Research Council (2009). *The Excellence in Research for Australia (ERA) Initiative*. Retrieved March 24, 2010, from http://www.arc.gov.au/era/

Bar-Ilan, J. (2008). Which h-index? — A comparison of WoS, Scopus and Google Scholar. *Scientometrics, 74*(2), 257-271.

Bates, R. (2003). Phelan's bibliometric analysis of the impact of Australian educational research. *Australian Educational Researcher, 30*(2), 57-64.

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. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation, 64*(1), 45-80.

Bornmann, L., & Daniel, H.-D. (2009). The state of h index research. *EMBO Rep, 10*(1), 2-6.

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H. D. (2008). Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics, 8*(1), 93-102.

Campbell, P. (2008). Escape from the impact factor. *Ethics in Science and Environmental Politics, 8*, 5-7.

Clarke, R. (2009). A Citation Analysis of Australian Information Systems Researchers: Towards a New ERA? *Australasian Journal of Information Systems, 15*(2), 23-44.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics, 69*(1), 131-152.

Goodyear, R. K., Brewer, D. J., Gallagher, K. S., Tracey, T. J. G., Claiborn, C. D., Lichtenberg, J. W., et al. (2009). The Intellectual Foundations of Education: Core Journals and Their Impacts on Scholarship and Practice. *Educational Researcher, 38*(9), 700-706.

Harzing, A. W. (2009). Publish or Perish (Version 2.8.3644).

Harzing, A. W. K., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics, 8*(1), 61-73.

Henzinger, M., Suñol, J., & Weber, I. (2010). The stability of the h-index. *Scientometrics, 84*(2), 465-479.

Herther, N. K. (2009). Research evaluation and citation analysis: key issues and implications *The Electronic Library, 27*(3), 361-375.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569-16572.

Hodder, A. P. W., & Hodder, C. (2010). Research culture and New Zealand's performance-based research fund: Some insights from bibliographic compilations of research outputs. *Scientometrics, 84*(3), 887-901.

Jarwal, S. D., Brion, A. M., & King, M. L. (2009). Measuring research quality using the journal impact factor, citations and 'Ranked Journals': blunt instruments or inspired metrics? *Journal of Higher Education Policy and Management, 31*(4), 289-300.

Levine-Clark, M., & Gil, E. (2009). A comparative analysis of social sciences citation tools. *Online Information Review, 33*(5), 986-996.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology, 58*(13), 2105-2125.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.

Moed, H. F. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics, 74*(1), 153-161.

Norris, M., & Oppenheim, C. (2010). Peer review and the h-index: Two studies. *Journal of Informetrics, 4*(3), 221-232.

Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics, 81*(3), 635-670.

Phelan, T. J., Anderson, D. S., & Bourke, P. (2000). Educational Research in Australia: A Bibliometric Analysis. In DETYA (Ed.), *The Impact of Educational Research* (pp. 575-671). Canberra: Higher Education Division Department of Education, Training and Youth Affairs.

Saad, G. (2010). Applying the h-index in exploring bibliometric properties of elite marketing scholars. *Scientometrics, 83*(2), 423-433.

Smith, A. (2008). Benchmarking Google Scholar with the New Zealand PBRF research assessment exercise. *Scientometrics, 74*(2), 309-316.

Vaughan, L., & Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics, 74*(2), 317-330.