The Spanish Journal of Psychology
2009, Vol. 12, No. 2, 833-849

ISSN 1138-7416

The Hirsch *h* Index in a Non-Mainstream Area: Methodology of the Behavioral Sciences in Spain

Miguel A. García-Pérez
Universidad Complutense (Spain)

The *h* index has advantages over journal impact factors for assessing the research performance of individuals, and it is becoming a reference tool for career assessment that is starting to be considered by some agencies as an aid in decisions for promotion, allocation, and funding. The *h* index has been reported to have adequate properties as a measure of the research accomplishments of individuals in areas where *h* values are usually high (i.e., at or above 40), but some concerns have been raised that its validity in other non-mainstream research areas is suspect. This paper presents data from an exhaustive computation and analysis of *h* indices for 204 faculty members in the area of Methodology of the Behavioral Sciences in Spain, an area where *h* indices tend to be low worldwide. The results indicate that the *h* index is substantially increased by self-citations and that the average *h* of full professors is not meaningfully larger than the average *h* of associate professors. Other interesting relations between *h* indices and demographic and academic variables are described, including the gender and age bias of *h*. In this field, but perhaps also in other fields where the average *h* is low, little justification is found for the use of the *h* index as a fair measure of research performance that can aid in funding or promotion decisions.
Keywords: citation analysis, self-citations, Hirsch index, h index, bibliometric indicators.

Frente a los índices de impacto, el índice *h* tiene ventajas para la evaluación de la carrera investigadora de personas individuales, y se está estableciendo como un índice de referencia en ese ámbito que empieza a utilizarse en la toma de decisiones acerca de promoción académica y concesión de ayudas para investigación. El índice *h* ha mostrado tener características adecuadas para esos propósitos en áreas en que los valores del índice son generalmente altos (40 o más), pero se han expresado dudas sobre su validez en áreas de menor saliencia. Este trabajo presenta un análisis exhaustivo del índice *h* de 204 profesores funcionarios del área de Metodología de las Ciencias del Comportamiento en España, un área en que el índice *h* suele tomar valores bajos mundialmente. Los resultados revelan que el índice *h* está significativamente inflado por autocitas, y que el índice *h* medio de los catedráticos del área no es sustancialmente superior al de los profesores titulares. El índice *h* también muestra otras relaciones interesantes con variables demográficas o académicas, incluyendo la presencia de sesgos ligados a la edad y el sexo. Todo esto sugiere que, en el área de Metodología pero quizá también en otras áreas, el uso del índice *h* en la toma de decisiones de promoción o financiación no está justificado.
Palabras clave: análisis de citas, autocitas, índice de Hirsch, índice h, bibliometría.

This research was supported by grant SEJ2005-00485 (Ministerio de Educación y Ciencia, Spain). I thank Jesús Salgado for his comments on an earlier draft of this paper.

Correspondence concerning this article should be addressed to Miguel A. García-Pérez, Departamento de Metodología, Facultad de Psicología, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, (Spain). Phone: +34-913943061; Fax: +34-913943189; E-mail: miguel@psi.ucm.es

In one way or another, scientific productivity has always been considered when making decisions about promotion or funding, and the use of objective indices for this purpose is welcome (García-Pérez, 2000). For a number of years, journal impact factors were regarded as an indicator of the quality of an individual's research output, but doubts have been raised about the extent to which this index yields a realistic portrait of the relative merits of applicants for academic placement and tenure (Colquhoun, 2003; Moed, 2002). It has indeed been shown that publication of a paper in a high-impact journal does not imply that the paper has been (or will be) influential (Seglen, 1997) and the position is now established that journal impact factors have no value in assessing individual scientists (Waheed, 2003). Journal impact factors are only an indicator of the salience of the journal in which a paper is published, and concerns have been aired to the effect that publishing in a high-impact journal is seen as more important than the content of the paper itself (Brookfield, 2003). A *Nature* editorial (*Nature*, 2005a) referred to the reliance on impact factors to assess the scientific quality of individuals as "unhealthy," and a further editorial (*Nature*, 2005b) rejoiced that a UK agency had prohibited assessment panels from judging papers by the impact factors of the journals in which they appeared.

In this context, a diametrically different approach, the *h* index (Hirsch, 2005), has swiftly gained popularity, perhaps because it considers directly the impact of each of the papers authored by an individual instead of that of the journals in which they came out. A scientist's *h* is the highest number of his/her papers that have each received at least that number of citations. Thus, a scientist with an *h* of 20 has 20 published papers each of which has received at least 20 citations. The *h* index was conceived and tested in disciplines where *h* tends to be high, and it has been shown to have predictive power in those areas (Bornmann & Daniel, 2005; Hirsch, 2007).

Appealing as it may seem, the *h* index is not without problems. To begin with, *h* was originally (and without proof) deemed robust to inflation by self-citations. This is probably correct in disciplines where *h* tends to be high and for the prominent scientists whose research records were the subject of the earlier studies, but it has later been shown that self-citations should definitely be excluded from computations at least in disciplines (and for individuals) where *h* tends to be moderate or low (Schreiber, 2007; Vinkler, 2007). Second, the *h* index is not the most precise and accurate single indicator of scientific quality (Kelly & Jennions, 2006, 2007; Lehmann, Jackson & Lautrup, 2007). Third, the *h* index has been shown to display gender and age effects (Kelly & Jennions, 2006, 2007), which questions the fairness of indiscriminate comparisons or rankings based on *h*. Fourth, the *h* index is biased in favor of individuals publishing papers with multiple co-authors (Batista, Campiteli, Kinouchi & Martinez, 2006). These and other shortcomings (see also Wendl, 2007) have led several authors to propose variants or generalizations of the *h* index (Batista et al., 2006; Bollen, Rodriguez & van de Sompel, 2006; García-Pérez, 2009; Ionannidis, 2008; Lehmann et al., 2007; Sidiropoulos, Katsaros & Manolopoulos, 2007; Taber, 2005), but so far the simplistic *h* index appears to be robust to evidence against itself.

Lining up with the enthusiasm which tends to consider the *h* index as the definitive scalar measure of research performance, and despite its potential and documented weaknesses, the funding agency of the Autonomous Government of Madrid (Spain) has recently issued an 80-page document (Rodríguez Navarro & Imperial Ródenas, 2007) setting standards on the use of *h* for institutional and individual assessments of research quality. The document includes instructions for the computation of *h*, and considers five different groups of disciplines according to variations in citation patterns and average journal impact factors. In the same vein, Salgado and Páez (2007) have explicitly advocated the use of the *h* index by agencies and committees responsible for awarding grants and making tenure decisions.

Although citation rates (the basis for the *h* index) have been shown to be unrelated to journal impact factors (Seglen, 1997)—which reveals that journal impact factors and *h* index do not measure the same construct—this lack of relation does not imply that the *h* index will indeed give the scalar measure that we aim at (on the assumption, of course, that we aim at summarizing a scientist's career by a scalar). The *h* index has been shown to display a variety of correlations (from low to high) with alternative indicators of research performance (van Raan, 2006). To the extent that these alternative indicators measure relevant aspects of scientific performance, low correlations indicate multidimensionality. Indeed, if *h* correlated very high with some other alternative indicator, what would the computation of *h* add to the picture provided by the alternative indicator? These moderate or low correlations, then, suggest that the *h* index should never be used in isolation.

As discussed above, the *h* index has been studied mostly in disciplines where its values are generally high, but its performance and dependability in other disciplines is largely unknown. The only study to that effect appears to be that of Salgado and Páez (2007), but their target population was limited to the small subset of full professors of Social Psychology in Spain. If *h* is to be used for promotion purposes (as Salgado and Páez advocated), the target population of study should rather be that of associate professors, the only ones who are yet to climb a further step along their academic career. And it is also useful to compare the *h* indices of full professors with those of the associate professors whose promotion is in their hands.

The goal of this paper is, thus, to compute and analyze the *h* indices of full and associate professors[1] of Methodology

[1] The Spanish terms for these tenured positions are, respectively, "catedrático" and "titular."

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

of the Behavioral Sciences in Spanish public universities, mostly with an eye towards assessing whether *h* indices might be useful for the purposes of grant awarding and academic promotion in a research field where available evidence (see Salgado & Páez, 2007, their Table 4) indicates that *h* indices are very low. A number of secondary goals include assessing the role of self-citations, checking for gender and academic-age bias, assessing correlations with other indicators of research performance, and assessing qualitatively the extent to which individuals with the same *h* can actually be regarded as having the same "global scientific impact" in the light of publication and citation practices in the area.

Parts of this work have been presented in abstract form (García-Pérez, 2008).

Method

Target Population

The target population for analysis consisted of 21 full professors (7 female and 14 male) and 126 associate professors (52 female and 74 male) at departments of Methodology of the Behavioral Sciences in Spanish public universities. These numbers do not exhaust the total number of tenured professors in those departments; they only represent the subset for which data were publicly available for being deemed eligible for membership in promotion committees[2] as of February 2007. A second (and potentially incomplete) list of non-eligible professors was separately compiled from different sources (mostly university web pages) that comprised 3 additional full professors (1 female and 2 male) and 54 additional associate professors (21 female and 33 male).

Data Collection

Publication and citation records were retrieved from the *ISI Web of Science* (WoS) on June 25–28, 2008, with all the precautions described elsewhere (García-Pérez, 2001). In particular, for an individual named Givenname Name1 Name2 at institution Y, the *Author Finder* utility of the WoS was used to retrieve publications authored by "Name1-Name2 G OR Name1Name2 G OR Name1 Name2 G OR Name1 G OR Name2 G N1" at institution Y (under any of its names or acronyms, selected from the list that the *Author Finder* utility provides). For individuals with two given names, the search template was also expanded to include each of the given names separately and in combination. The case of individuals with more complex names (e.g., De la Serna) was treated along the same lines and resulted in even more expanded search templates.[3]

Because of the well-known mis-citations, mis-spellings, and errors of various sorts committed in the compilation of the WoS data base (Brown, 1999; Kotiaho, 1999; Kotiaho, Tomkins & Simmons, 1999; Price, 1998), missing or misplaced records in the WoS were sought by comparing the results of this search against the actual CV of 4 full professors and 24 associate professors in our main list of eligible professors (19% of the 147 individuals). Of course, this comparison can only reveal errors in publication records (but not in citation records) retrieved from the WoS for this subset of test individuals, and it cannot serve but as a gross indicator of the accuracy and completeness of the publication information retrieved from the WoS. If the CV of some professor included qualifying[4] papers that our WoS search had not located, a direct search for each of those papers in the WoS was carried out and, if successful, the data file was updated using the newly retrieved information. This procedure mimics what each individual researcher would actually do in case the initial WoS search for citation information on his/her own papers did not retrieve all of them, and helps gathering all the information that is actually available (although perhaps misplaced or mis-filed) in the WoS. If a record for a qualifying paper was not found in the WoS as a result of this direct search, no modification to the data file was made.[5]

---

[2] The Spanish referent here is "tribunales de habilitación."

[3] Salgado and Páez (2007) suggested an alternative search strategy still within the WoS which consists of searching by author under the *Cited Reference Search* utility. We tried out and then abandoned this strategy because in many cases (in particular, for authors with relatively common last names) the resultant list of publications with their accompanying citation records could not be reasonably attributed unmistakably to the incumbent authors. This strategy would then have unfairly inflated the publication and citation records of authors with common names, thus discriminating against authors with less common names. It would also have produced unrealistically inflated publication and citation counts.

[4] The CVs generally included papers that could not possibly be found in the WoS because the journals (or the dates) in which they were published are not covered in the data base. More often than not, these were pre-1990 papers written in Spanish (or in some of the other co-official languages in Spain) and published in Spanish journals. Hence, the impact of these papers on the global scientific community (had this information been available in the WoS data base) is likely to have been almost null. These papers are thus considered as non-qualifying, and the term "qualifying paper" refers instead to papers that should have been retrieved in our WoS search because the journals and dates in which they appeared are actually covered in the WoS.

[5] The reason for this is that several errors were observed across the 28 CVs that were handled, not only in volume or issue numbers, page ranges, or publication dates of the papers listed on the CVs, but also in the journal names themselves. Determining what type of error each of these were was sometimes impossible and, in those cases, the paper was not even counted as one more publication by the corresponding individual because it was not clear that it would have been included in the WoS.

In comparison with Salgado and Páez (2007), no further data bases were used to complement this information, for several reasons. First, many journals that were not included in the WoS at the time of the analysis of Salgado and Páez have been included in the elapsed time. Second, it was found that APA's *PsycINFO* data base rendered shorter publication lists and lower citation counts than those obtained from the WoS for a set of test individuals, despite the fact that the former includes some (but, definitely, not the majority of) book chapters and Spanish psychology journals (whose consideration did not contribute nevertheless to the *h* of these test individuals). Third, the Spanish data base *In-Recs* (used by Salgado and Páez) includes journals that are not suitable referents for a global *h* index. Our goal in this paper is to investigate the performance of the *h* index at what it is purported to be, namely, a measure of global (i.e., worldwide) impact and research performance that is not contaminated by the inclusion of research coming out in publication black holes. It is nevertheless unlikely that the papers that were excluded by not incorporating (less decisive) information from alternative data bases will have differential effects that might invalidate the results of our analysis.

Computation of the h Index and Other Performance Measures

The search procedure described in the preceding section rendered, for each individual, a list of his/her publications which was sorted by "times cited" using the WoS facility.[6] The first analysis consisted of trimming the list by removing non-qualifying publications such as obituaries, biographical notes, editorial comments, introductions to special issues, meeting abstracts, errata, book reviews, etc. It should be noted that these items hardly contribute to an author's *h* because they are rarely cited more than once or twice,[7] but they certainly inflate the author's count of publications. Next, three indices were computed: (1) the total number N_p of papers in the trimmed list, (2) the conventional *h* as it turns out from the trimmed list, and (3) a corrected *h*, dubbed here h_c, which is the value of *h* when self-citations are excluded. A self-citation is defined here as a citing paper one of whose co-authors matches some of the co-authors of the target paper. In other words, the corrected citation count for a target paper is the number of citing papers published by authors not including any of the authors of the target paper. The trimmed list was stored on file for subsequent analyses and, thus, other indices such as the total number of citations and the maximal number of citations were also available.

Results

Accuracy and Completeness of the Retrieved Information

The search for potentially missing or misplaced records in the WoS through a comparison with the actual CV of 28 individuals yielded three different situations:

(1) Perfect agreement, defined as there being a record in the WoS for each and all of the qualifying papers on the CV (11 of 28 cases; 39.3%);

(2) missing papers on the CV, either because the WoS search went further back in time than the CV (2 of 28 cases; 7.1%), because the WoS search retrieved fresh-published papers that were not mentioned in the CV (4 of 28 cases; 14.3%) or, more often, because of the omission of papers in the CV that were nevertheless published during the period that it nominally covered (6 of 28 cases; 21.4%); and

(3) missing papers on the WoS list (13 of 28 cases; 46.4%). The latter errors were often attributable to inappropriate (as it turned out) decisions when refining the huge initial lists of records retrieved for individuals with extremely common names (e.g., Fernández, García, Gómez) who publish under their given name and first last name only, but in some other cases the error was caused by the name of the author being mis-spelled or mistaken in the WoS data base, or by a record for the paper being completely missing in the WoS (when this additional search was actually carried out by journal name, volume, issue, and page numbers).

As discussed in the Methods section, citation information for the papers that could be found in the WoS only with the help of the author's CV was nevertheless used to update our data file, but its consequences were minimal: Publication counts certainly increased minimally for some individuals with this additional information, but the number of citations and the *h* index did not generally vary except for three individuals whose *h* was low to begin with. In sum, the WoS appears to be a reliable (though not perfect) source of publication data, and it is unlikely that our results are meaningfully affected (or biased) by potentially missing records in the WoS for individuals whose CV was not requested so as to carry out the comparison whose results have just been described.

Distributions of h and h_c

Figure 1 shows histograms and descriptive statistics of the conventional *h* index separately for the 21 eligible

[6] The resultant data files are available from the author upon request.

[7] However, these meager citation counts can influence the *h* of a non-negligible number of individuals with $h < 3$.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

full professors (Fig. 1a), the 126 eligible associate professors (Fig. 1b), the 3 non-eligible full professors (Fig. 1c), the 54 non-eligible associate professors (Fig. 1d), and also for aggregated data from all 204 professors (Fig. 1e). Besides the overall low values of h and their low average, the most significant aspect is that the distributions for all four groups overlap almost completely, with two eligible associate professors (Fig. 1b) having h indices above the range of values found for eligible full professors (Fig. 1a). It is also clear by eye that non-eligible professors have lower average h and narrower ranges than their eligible colleagues (compare Fig. 1a with Fig. 1c and Fig. 1b with Fig. 1d), on the caveat that the number of non-eligible full professors is certainly too small to warrant the comparison. As for eligible professors (Figs. 1a and 1b), an homogeneity test whose details are described in the Appendix revealed that the distributions of h among full and associate professors were not significantly different, indicating that the groups of eligible full professors and eligible associate professors cannot be distinguished by the distribution of their h indices.

For the entire set of 204 professors, h values ranged from 0 to 9 (excluding an outlier whose h is 15) with an average of 2.06 and a standard deviation of 2.04, which thus represents the characteristics of the h index in the field of Methodology of the Behavioral Sciences in Spain.

Figure 2 shows analogous results for the corrected index h_c, which excludes self-citations from the computation of h. Quite clearly, the average h_c is smaller than the average h in all groups. For aggregated data from the 204 professors (Fig. 2e), h_c values ranged from 0 to 7 (excluding again an outlier whose h_c is 12) with an average of 1.67 and a standard deviation of 1.67, representing an average drop of 0.4 units when self-citations are excluded. To further appreciate the differences between h and h_c at the individual level, Fig. 3 shows a tabulated scatterplot for the set of 204 professors. Each cell in the plot corresponds to a particular pair of values for h (along the abscissa) and h_c (along the ordinate), with inset numerals indicating the number of individuals with that pair of values. Compared to the value of h, h_c drops by three units for 3 individuals, by two units for 10 individuals, by a unit for 52 individuals, and remains invariant for 139 individuals.

Gender Bias

Figure 4a shows histograms of the distribution of h (left panel) and h_c (right panel) for female professors and Fig. 4b similarly shows histograms for male professors. As is clear from panels in the left column of Fig. 4, the average h of male professors is 0.7 units higher than that of female professors, whose h hardly ever exceeds 3. By an unconditional separate-variance t test (for justification, see Hayes & Cai, 2007; Zimmerman, 2004), this difference is statistically significant ($t_{202} = 2.65$, $p < .01$). The same holds literally for the corrected index h_c (right column of Fig. 4), which on average is about 0.4 units smaller than h in each group.

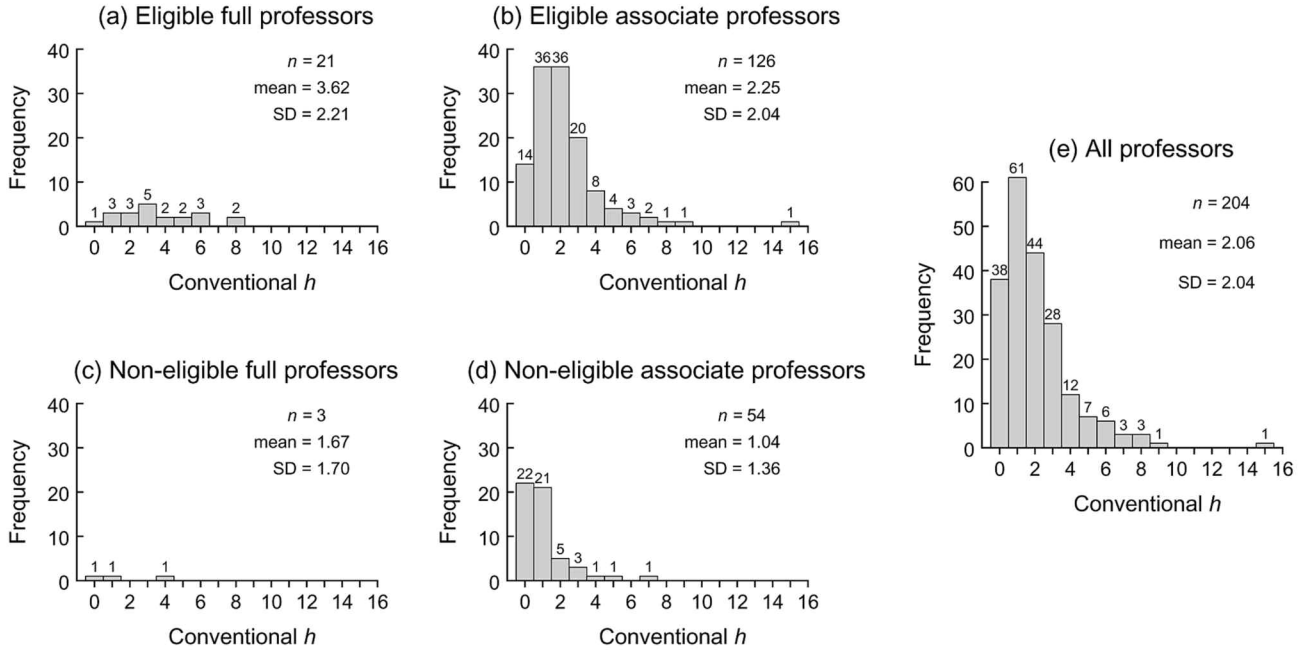


Figure 1. Distribution of the h index among eligible full professors (a), eligible associate professors (b), non-eligible full professors (c), non-eligible associate professors (d), and for the total sample of professors. Sample size, mean, and standard deviation (SD) are shown in the insets in each case. The numeral at the top of each bar indicates the number of cases.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

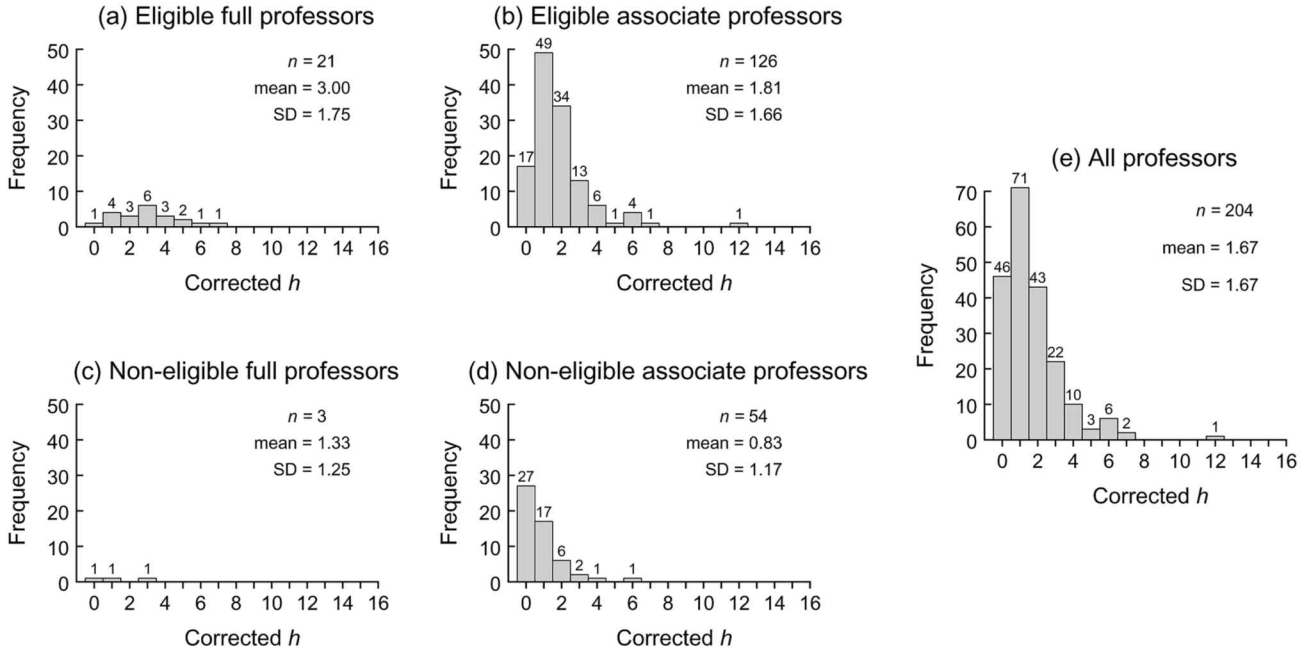


Figure 2. Distribution of the corrected index h_c among eligible full professors (a), eligible associate professors (b), non-eligible full professors (c), non-eligible associate professors (d), and for the total sample of professors. Sample size, mean, and standard deviation (SD) are shown in the insets in each case. The numeral at the top of each bar indicates the number of cases.

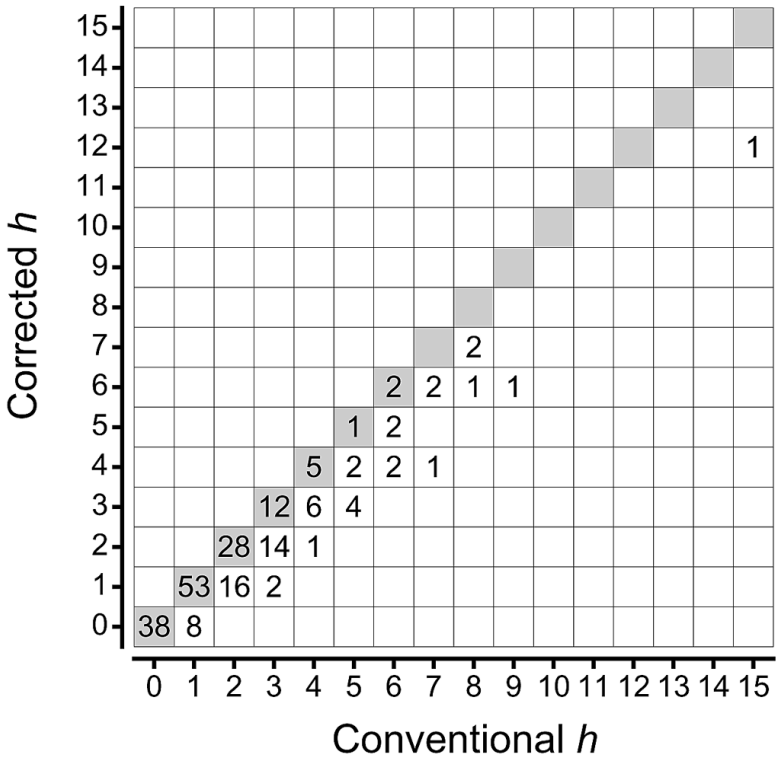


Figure 3. Tabulated scatterplot of the corrected index h_c against conventional h for the total sample of 204 professors. Equality of h and h_c occurs along the diagonal (grayed cells).

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

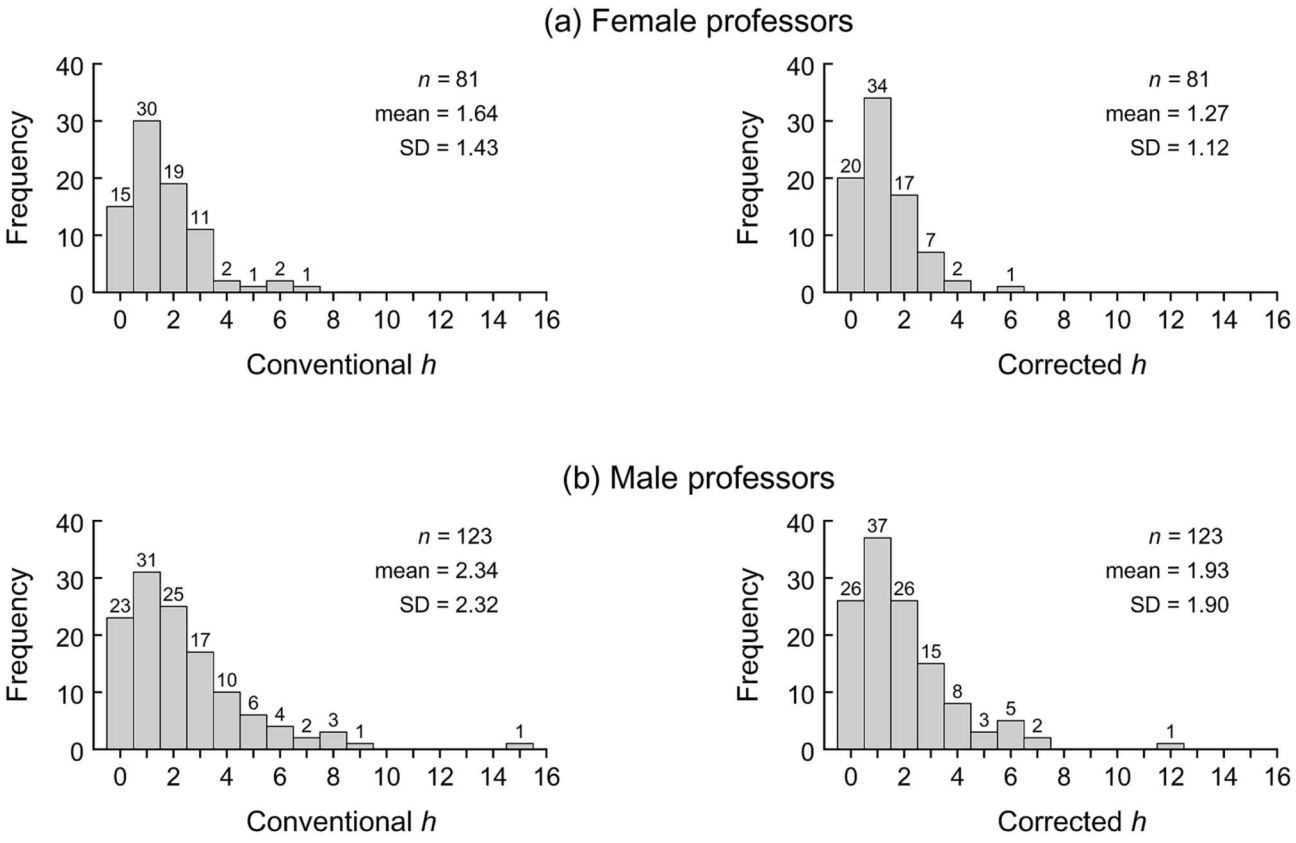


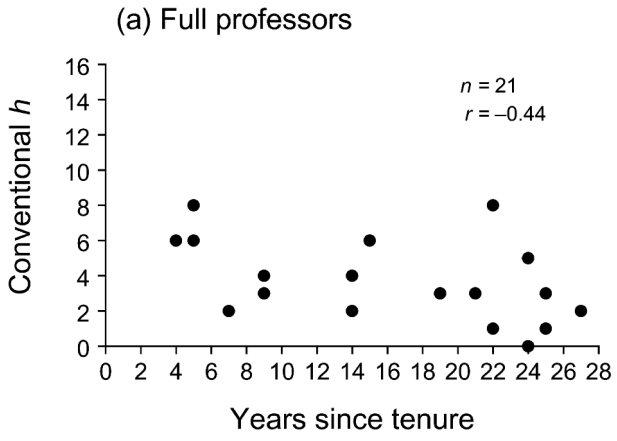
Figure 4. Distribution of the *h* index (left) and the corrected index h_c (right) among female professors (a) and male professors (b). Sample size, mean, and standard deviation (SD) are shown in the insets in each case. The numeral at the top of each bar indicates the number of cases.

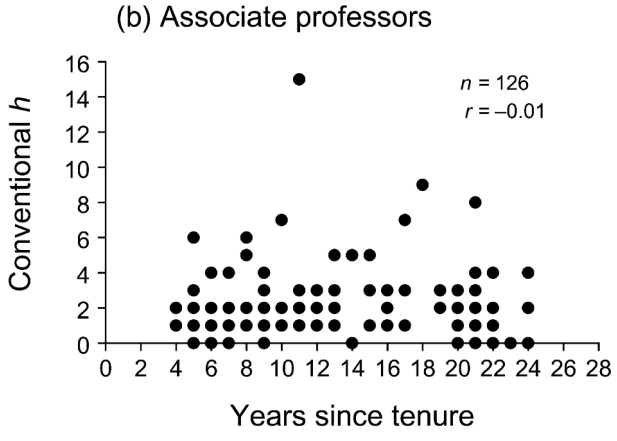
Academic Age Bias

Figure 5 shows the relationship between academic age (measured in number of years since tenure) and *h*, separately for eligible full professors (Fig. 5a), eligible associate professors (Fig. 5b), and aggregated data from all eligible professors (Fig. 5c). This analysis is restricted to the groups of eligible professors because the year of tenure of non-eligible professors was not available. Interestingly, the product-moment correlation is negative and statistically significant ($r = –.44$; $p < .025$) for full professors, whereas it is non-significantly different from zero for associate professors ($r = –.01$) or for aggregated data ($r = –.04$). These results are at odds with positive and statistically significant correlations reported in other fields (e.g., Kelly & Jennions, 2006), although a negative correlation was also reported by Salgado and Páez (2007) in the case of full professors of Social Psychology in Spain. The relationship between academic age and corrected *h* (not shown graphically) was similar, yielding product-moment correlations of –.40 for full professors, –.01 for associate professors, and –.02 for aggregated data from all professors.

It can be argued that number of years since tenure is not a reliable indicator of academic age, if only because it underestimates the academic age of full professors who obtained an associate professorship years earlier. An indicator that is free of this differential bias is the number of years since each individual published his/her first paper in a journal covered in the WoS, the indicator that Hirsch (2005) actually used. These data are obviously available for all individuals in our sample in the files that our WoS search retrieved. The relation between academic age measured in this alternative way and *h* is shown in Fig. 6, separately for all full professors (Fig. 6a), all associate professors (Fig. 6b), and aggregated data from all professors (Fig. 6c). Note that the number of associate professors in Fig. 6b is only 163, and this is because 5 eligible and 12 non-eligible associate professors have not published any paper in the journals included in the WoS. With this alternative indicator, the relation between academic age and *h* is positive although somewhat lower than that found in other fields (Kelly & Jennions, 2006), and it is similar for full and associate professors.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press





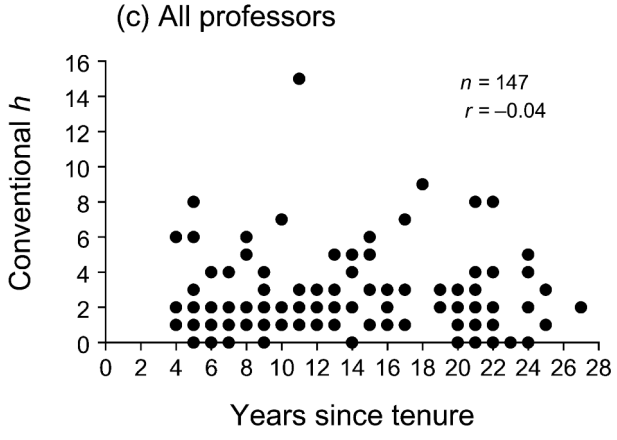
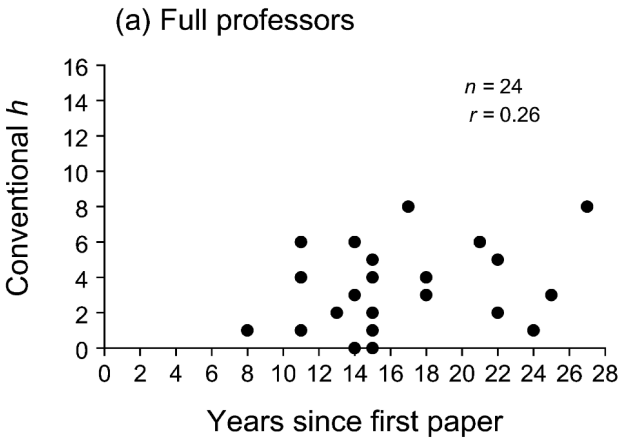
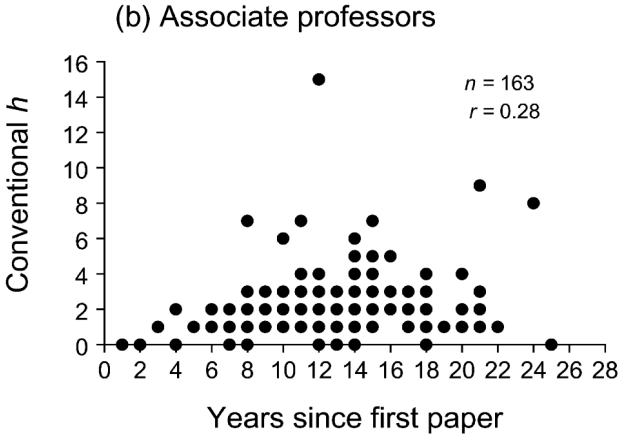


Figure 5. Relationship between the *h* index and academic age (number of years since tenure) among eligible full professors (a), eligible associate professors (b), and for the total sample of eligible professors (c). Sample size and product-moment correlation (*r*) are shown in the insets in each case.





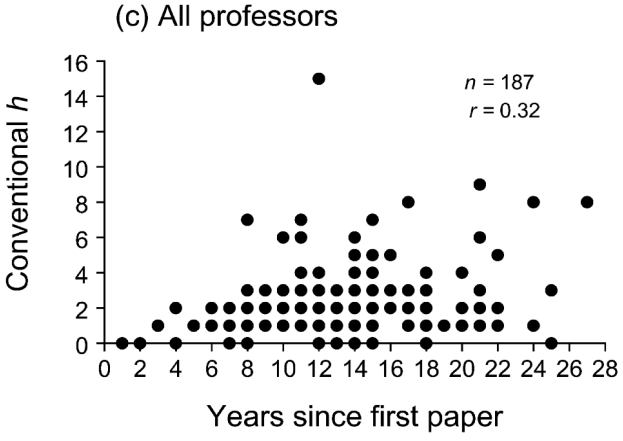


Figure 6. Relationship between the *h* index and academic age (number of years since first paper) among all full professors (a), all associate professors (b), and for the total sample of professors (c). Sample size and product-moment correlation (*r*) are shown in the insets in each case.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

Relation to Other Indices

Justification for the use of h arises if it is not very highly correlated with alternative indicators of research performance. Otherwise, h would not provide any additional information that is not already carried by the alternative indicator. The first such indicator is the total number of citations, N_c, received across all the N_p papers published by an individual. The lower bound for N_c is h^2, because each of the contributing h papers must have received at least h citations. Generally, however, some of those h papers will have received more citations than this minimum, and the individual will also have published other papers which may have also been cited. Hirsch (2005) proposed the relation

$$N_c = a\, h^2, \tag{1}$$

where $a = 1$ would indicate that all individuals in the field attain their h with the minimum requirements for citations whereas higher values for a would occur either when contributing papers receive increasingly higher numbers of citations than necessary or when individuals publish increasing

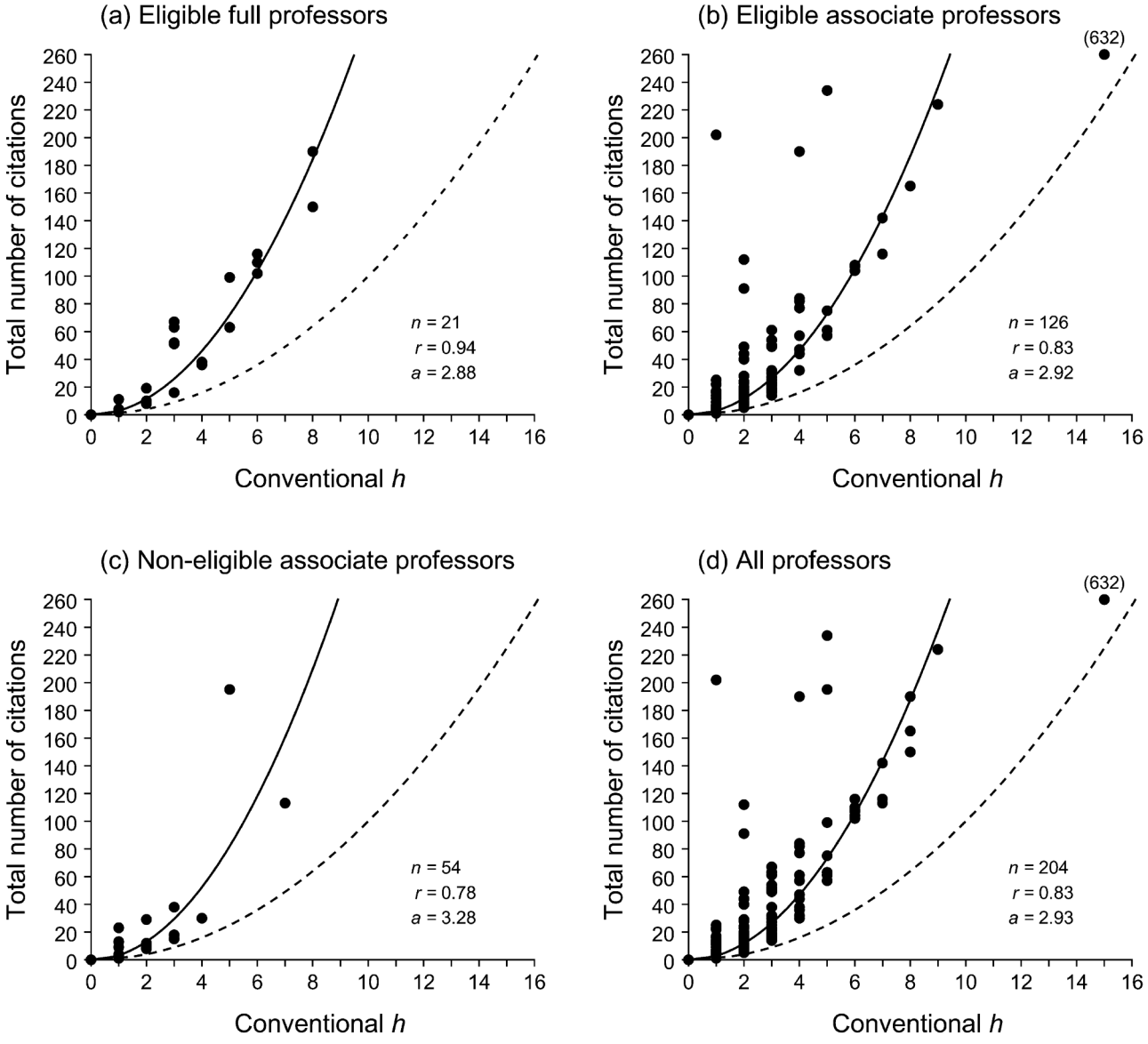


Figure 7. Relationship between total number of citations and conventional h among eligible full professors (a), eligible associate professors (b), non-eligible associate professors (c), and for the total sample of 204 professors including also three non-eligible full professors (d). Sample size, product-moment correlation (r), and estimated a in Eqn. (1) are shown in the insets in each case. The solid curve in each panel is the best-fitting quadratic function in Eqn. (1) with the estimated a in each case; the dashed curve indicates the lower bound for the relation. The number in parenthesis near the data point at the top right in panels (b) and (d) is the actual ordinate for that point.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

numbers of non-contributing papers that are also cited. For our data, the relations of N_c to h for eligible full professors, eligible associate professors, non-eligible associate professors, and all professors are shown, respectively, in Figs. 7a, 7b, 7c, and 7d. The dashed curve in each panel is the lower bound given by $N_c = h^2$ and the solid curve is the least-squares fit of Eqn. (1) to the data in each panel, yielding the estimates of a shown in the insets. Also reported in the panels of Fig. 7 are the product-moment correlations between h and N_c.

The values for a reported across the panels of Fig. 7 for each group of professors and for the aggregate are in reasonable agreement with those reported by Hirsch (2005) in mainstream fields and also with those reported by Salgado and Páez (2007) for full professors of Social Psychology in Spain. Note also that the correlation between h and N_c is generally high despite the quadratic relation implied by Eqn. (1), although the correlation is slightly lower in Figs. 7b, 7c, and 7d due to the presence of individuals with relatively low h but with a large total number of citations. Inspection of the data file revealed that the total citation count for these individuals came mostly from citations received by just one of their papers. Removal of the individuals with these unusual patterns resulted in correlations of .88 in Figs. 7b, 7c, and 7d, and also brought together the estimates of a to a common value of 2.9 in all panels. These correlations are similar to those reported by Salgado and Páez (2007) for Spanish full professors of Social Psychology.

When the corrected index h_c was used instead of the conventional h, estimated values for the scale factor a in Eqn. (1) raised to about 4 for each group and for the aggregate, although correlations remained very similar to those reported in the preceding paragraph for the case of conventional h. It should be noted that the total number of citations for this analysis also included self-citations, which were thus only removed from consideration for the computation of the corrected index h_c but not for the computation of overall number of citations.

A second alternative indicator of the impact of an individual's research is the largest citation number, that is, the largest number of citations received by one of his/her individual papers. Salgado and Páez (2007) reported a correlation of .61 between h and largest citation number for Spanish full professors of Social Psychology. Figure 8 shows the relationship between these indicators for aggregated data from all 204 professors in our sample, which attains a correlation of only .41. The correlations were similar in each of the separate groups of professors, and this is the reason that results are reported only for the aggregate. The reason for this comparatively low value is, again, the presence of a few individuals with low h but relatively high (i.e., above 80) largest citation number. These are the same individuals involved in the situation described above with respect to Fig. 7, and removal of their data resulted in a correlation of .76. Finally, the relationship between corrected h_c and largest citation number was similar both in the pattern displayed by the data and in the magnitude of correlation.

Finally, the relationship between h and total number of published papers was approximately linear and the correlation was 0.82, as Fig. 9 reveals for aggregated data from all professors. The picture was similar in each of the

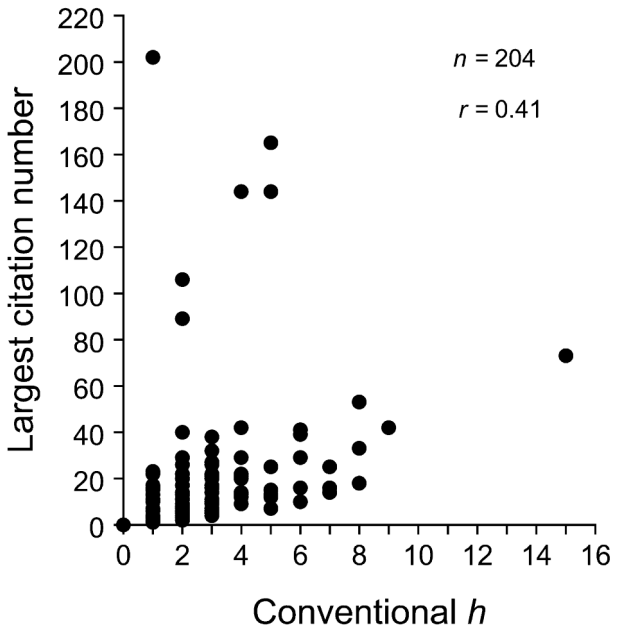


Figure 8. Relationship between largest citation number and conventional h for the total sample of professors. Sample size and product-moment correlation (r) are shown in the inset.

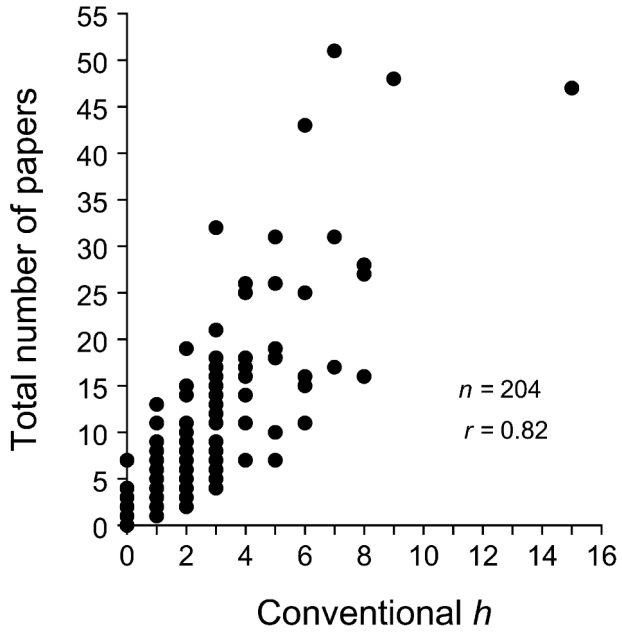


Figure 9. Relationship between total number of papers and conventional h for the total sample of professors. Sample size and product-moment correlation (r) are shown in the inset.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

separate groups of professors, and indicates that the total number of papers published in journals covered in the WoS data base is a good predictor of h. The picture was very similar for the corrected index h_c.

Quantitative and Qualitative Comparison of the Individuals With the Higher h

With a sample of 204 individuals and only 11 distinct h values, ties abound (see Fig. 1). It is nevertheless very unlikely that individuals with the same h will have comparable research careers. To look into this issue, a first quantitative analysis of the top 14 individuals[8] as regards their h was carried out whose results are presented in Table 1. Individuals are listed left to right by decreasing h, and individuals with the same h are listed left to right by decreasing total number of citations.

Consider the six individuals with $h = 6$ (rightmost six columns in Table 1). Papers published by all of these individuals have received a similar total number of citations (between 102 and 116) but the total number of papers published varies between 11 and 43, and the number of papers published in the last five years varies between 3 and 24. In four of these six individuals, the total citation count for their six contributing papers is almost identical to the total number of citations received by all of their papers, revealing that their non-contributing papers (which vary in number between 5 and 19) have barely received any citation. The remaining two individuals have published 43 papers each (substantially more than any of the four individuals just referred to), but recall that only six of these papers have received six or more citations each (by necessity, since $h = 6$ for these individuals); this characteristic indicates a relatively vast production that has not received any attention by the global scientific community. Finally, it is also interesting to note that the median number of authors across the six contributing papers varies for these individuals between two and four, with a minimum of two and a maximum of eight.

Similar quantitative comparisons can be made for individuals sharing other values of h, or for individuals whose h differs only by a small amount. For instance, the first of the three individuals with $h = 8$ has published a total of 27 papers (8 in the last five years), whereas the first of the three individuals with $h = 7$ (the next lower value) has published a total of 51 papers (23 in the last five years); it is also interesting to note that the first of these two individuals published his first paper 12 years before the second

Table 1
Quantitative comparison of the 14 individuals with the largest h

h	15	9	8	8	8	7	7	7	6	6	6	6	6	6
Group[a]	EA	EA	EF	EA	EF	EA	EA	NA	EF	EF	EA	EA	EA	EF
Year of tenure	1997	1990	1986	1987	2003	1991	1998	2008	2004	1993	2003	2000	2000	2003
All papers														
Year of first paper	1996	1987	1981	1984	1991	1993	1997	2000	1997	1994	1994	1998	1998	1987
Year of last paper	2008	2007	2007	2007	2008	2008	2008	2007	2008	2007	2007	2007	2006	2008
Total no. papers	47	48	27	16	28	51	31	17	25	43	16	15	11	43
No. papers in 2003–2007	26	16	8	4	10	23	16	11	5	24	5	7	3	13
Total no. citations	632	224	190	165	150	142	116	113	116	110	108	107	104	102
Contributing papers														
Total citation count	502	127	153	154	104	72	79	90	99	56	93	97[b]	97[b]	47
Largest no. citations	73	42	53	33	18	14	25	16	29	16	39	41[b]	41[b]	10
Second largest no. citations	57	13	42	32	16	12	15	16	22	11	19	17[b]	17[b]	9
Smallest no. authors	1	1	3	1	2	1	2	4	2	2	2	2[b]	2[b]	2
Largest no. authors	3	2	7	4	4	2	6	6	5	8	3	4[b]	4[b]	4
Median no. authors	2	2	5	2.5	3	1	3	4	3	4	2.5	3.5[b]	3.5[b]	2

[a] EA: Eligible associate professor; EF: Eligible full professor; NA: Non-eligible associate professor.
[b] These two individuals are co-authors in the six papers that contribute to their h, hence the equality of all these values.

[8] The reason for picking 14 rather than the typical 10 individuals is that there are 8 individuals with $h > 6$ and 6 individuals with $h = 6$.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

individual, which gave him 80% more time to publish; on another dimension, the number of authors in the contributing papers published by the first of these two individuals varies between three and seven with a median of five, whereas the number of authors in the contributing papers published by the second of these individuals varies between one and two with a median of one.

These and other comparisons confirm that *h* is a very poor summary of quite diverse research careers. Furthermore, the impact of the research described in a given paper cannot be definitely attributed to any of the co-authors in multi-authored papers. These judgments require qualitative analyses that scrutinize the production of each author in search for solid and stable lines of research as well as for continued collaborations with other authors. It should not be overlooked that the *h* index aims at measuring the impact of an individual's contribution, which in turn implies that the individual has an identifiable line of research (which most of his/her publications address and which defines the scientific community where the individual makes his/her impact), and that the individual perhaps has an identifiable network of co-authors (besides, of course, sporadic co-authorship with, e.g., Ph.D. students).

The publication records of the individuals for whom quantitative data are listed in Table 1 were thoroughly analyzed as described in the preceding paragraph in search for evidence that the individuals' *h* is indeed the measure of impact that is intended. This analysis revealed two quite different scenarios. In most cases (12 of 14), active and influential research lines were clearly identifiable and papers contributing to the *h* of these individuals were among the set of papers defining their research line. In the two other cases a research line was also clearly identifiable but, in contrast, contributions to that research field had not received enough citations to back up the individual's *h*: Most of the papers contributing to the *h* of these individuals come instead from a diversity of fields (not even in Psychology in many cases) and have each a long list of co-authors among whom these individuals appear to have merely participated as data analysts. Then, the incidental *h* of these two individuals can hardly be taken as a true measure of the impact of their own research.

Finally, it is interesting at this point to compare how the standings of the top 10 full professors of Methodology identified by Salgado and Páez (2007; their Table 4) have changed in the elapsed time. It should be noted that this comparison may only reflect the differences between the stringent criterion used in this study (reliance on data retrieved only from the WoS) compared to the more indulgent criterion of Salgado and Páez (which also included data retrieved from *InRecs*, a data base for research published in Spanish). Actually, three of the top ten individuals listed by Salgado and Páez are not among the top ten full professors according to our records, and the discrepancy for these individuals is substantial both in terms of their *h* index and in terms of the largest citation count (see Fig. 10).

Most-cited papers in the Field

Table 2 lists the titles of the 11 most-cited papers among those published by individuals in our sample, giving also the total number of citations that each paper has received, the total number of authors in that paper, the position of the incumbent author(s) within the list of authors, and the *h* of the incumbent author(s). It is noteworthy that the topic of the most cited paper is totally alien to Psychology, further corroborating that professors of Methodology often help out with the analysis of data in other disciplines (see also the paper ranking in 9th place in the list of Table 2). It is also noteworthy that the authors of these most-cited papers do not generally have the largest *h* among the 204 individuals in our sample, further indicating the frailty of a mere citation count and its casual relation to the impact of the overall research of the author.

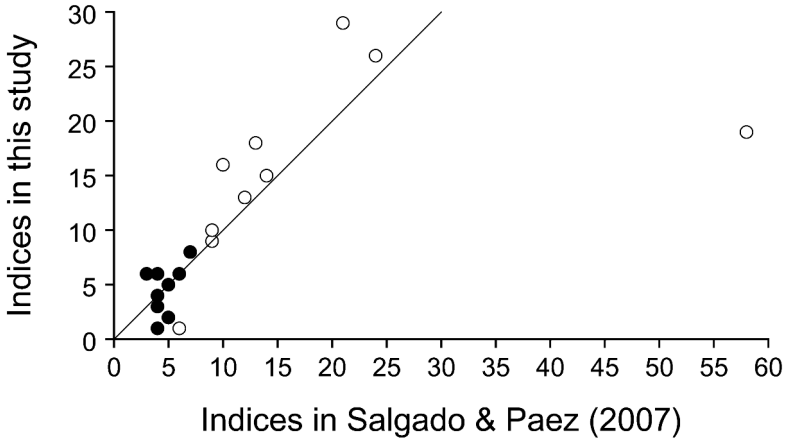


Figure 10. Indices reported by Salgado and Páez (2007) for the top 10 professors of Methodology (abscissa) compared to the same indices obtained in this study (ordinate). Solid symbols indicate the *h* index; open symbols indicate the largest citation count.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

Table 2
Most-cited papers

Title	NC[a]	NA[b]	PIA[c]	*h*[d]
1. Bone disease in predialysis, hemodialysis, and CAPD patients – Evidence of a better bone response to PTH	202	14	9	1
2. Cerebral lateralization of language in normal left-handed people studied by functional MRI	165	4	3	5
3. Sistema de clasificación del método en los informes de investigación en Psicología[e]	144	2	1,2	5,5
4. International preferences in selecting mates: A study of 37 cultures	106	50	6	2
5. Vascular dementia – A clinicopathological study	89	6	4	2
6. Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons	73	3	2	15
7. Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming	57	2	1	15
8. Effects of syllable frequency and syllable neighborhood frequency in visual word recognition	55	2	1	15
9. Prediction of therapeutic failure in patients with bleeding peptic ulcer treated with endoscopic injection	53	7	4	8
10. Diagnostic agreement between clinicians and the diagnostic interview for children and adolescents – DICA-R – in an outpatient sample	42	6	3,4,5	8,5,4
11. Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties	42	1	1	9

[a] NC: Number of citations
[b] NA: Number of authors
[c] PIA: Position of the incumbent author(s). Multiple elements indicate that several authors are individuals in our sample
[d] *h*: *h* index of the incumbent author(s). Multiple elements indicate the *h* of each of several authors who are individuals in our sample
[e] This paper is actually written in Spanish and has no title in English

Discussion

The work described in this paper mainly aimed at determining the characteristics of the *h* index in the field of Methodology of the Behavioral Sciences in Spain, but also at assessing the extent to which these characteristics would justify the use of *h* indices as an aid for grant awarding and decisions about academic placement and tenure. The results have confirmed that *h* is very low in this field, that the index is somewhat inflated by self-citations, that it shows gender bias, and that its characteristics make it non-dependable as an indicator of research performance. The following subsections elaborate on these statements.

Relation of h to Academic Position and Committee Eligibility

Results presented in Figs. 1 and 2 reveal that neither the *h* index nor the corrected index h_c are meaningfully related to academic position. The range and overlap of the distributions for full professors is similar to that for associate professors and, if anything, there are two associate professors whose *h* is higher than that of the full professors with the higher *h* (this remains true only for one associate professor when the corrected index h_c is considered).

As for the relation of *h* (or h_c) with committee eligibility, it is also clear from Figs. 1 and 2 that non-eligible professors have lower average *h* and h_c than their eligible counterparts, and that the ranges of the distributions have a lower upper limit in the case of non-eligible professors. Nevertheless, eligible full and associate professors can be found whose *h* or h_c indices are similarly as low-valued as those of non-eligible professors. In other words, *h* and h_c indices cannot differentiate eligible from non-eligible professors either. To the extent that eligibility for promotion committees in Spain is presumably based on research performance, the *h* index does not appear to capture the relevant aspects of research performance.

The Value of Citations

The logic underlying the design of the *h* index is that authors whose contribution is relevant (as judged by their peers) are cited more often than authors whose contribution is less decisive for the progress of a field. In a field where *h* tends to be high (at or above, say, 50), the *h* index is

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

likely to be a dependable measure of the relevance of an individual's contribution because these high *h* values are unlikely to be inflated by spurious factors such as self-citations or by the intrusion of papers that do not really represent a contribution of the incumbent author. But this is unlikely to hold in fields where *h* indices are low.

On the other hand, citations (and, hence, *h*) do not always reflect the actual contribution of the incumbent author: In our target population, two individuals with an *h* in the top quartile have published papers in their main area of research that have received very few citations whereas, at the same time, these individuals often participate as data analysts in multi-authored papers in alien fields, which is where they get their larger citation counts and, hence, their *h*. A mere count of citations does not take into account the extent to which a given paper represents a significant contribution by each of the co-authors.

On the other hand, papers that are widely considered influential in their field may have low citation counts. This situation is often caused by the authors' demonstration of a fundamental flaw that contributed to extinguishing some line of research. A well-known case in point is Yellott's (1969) demonstration that the basic assumptions underlying the stochastic learning models that were trendy in the 1950s and 1960s were fundamentally flawed (see Falmagne, 2005): As of July 16, 2008, the citation count of Yellott's paper in the WoS is a mere 33 despite its tremendous influence in discontinuing research aimed at elucidating the question of whether humans set for probability matching or for probability maximizing (i.e., whether they learn to match choice probabilities to outcome probabilities or, rather, they learn to make the choices that maximize their success rate).

Finally, reliance on *h* seems to be triggering citation coalitions and establishing citation practices that further pervert the ultimate goal of *h* (see Todd & Ladle, 2008).

The Role of Self-Citations

Self-citations are not necessarily evil. Indeed, they are a must for active scientists with a well-established line of research and whose current research questions build upon their own earlier results as much as they do on the work of others who also cite them. However, self-citations inflate *h* indices undeservedly when hardly any other author cites one's papers. In a sense, an adequate number of self-citations indicates a steady line of research, but the worldwide relevance of this line of research is suspect when self-citations are not accompanied by a number of citations by other authors. It should be borne in mind that the *h* index was designed to assess the global impact of a researcher's contribution, which seems to subsume the notions of recognition by other authors and influence on their work.

Publication and Citation Practices

Out of the 402 papers contributing to *h* indices across our main sample of 21 eligible full professors and 126 eligible associate professors, 98 (24.4%) had come out in *Psicothema*, and 85 of those 98 (86.7%) were written in Spanish. Although these publications are legitimate contributions to *h* in a strict sense, their global status is certainly questionable in terms of the worldwide availability of the contents of the papers (as opposed to the accessibility of the papers themselves, which are freely available at http://www.psicothema.com).[9]

The peculiar status of *Psicothema* was pointed out in an earlier analysis of published research by Spanish professors in the area of Methodology of the Behavioral Sciences (García-Pérez, 2001), and it does not seem to have changed noticeably a decade down the road. For instance, Musi-Lechuga, Olivas-Ávila, Portillo-Reyes and Villalobos-Galvis (2005) reported that 20.88% of the production of Spanish tenured professors in Psychology[10] is to be found in the pages of *Psicothema*: 1,370 papers altogether, compared to 197 papers in the journal ranked in second place. For the case of professors in the area of Methodology, the total number of papers published in *Psicothema* was 423, with 42 papers coming out in the journal ranked second. As of July 1, 2008, a WoS retrieval of papers published in *Psicothema* rendered a total of 1,626 records and indicated a journal *h* of 12 and a corrected h_c of 8 (computed by removing citations coming from *Psicothema* itself; the corrected h_c was 10 when computed by removing citations coming from the target authors); when sorted by number of times cited, the number of citations received by the top *Psicothema* paper was 30, and the top ten papers received an overall number of 200 citations 96 of which (48%) came from *Psicothema* itself. These citation practices certainly boost the journal impact factor and *h* index of *Psicothema*, but the fact that most of the papers published in *Psicothema* are written in Spanish implies that their actual impact on the global scientific community is less than the value of *h* might imply.

---

[9] This statement should not be misconstrued. The scientific quality and relevance of the papers coming out in *Psicothema* were not analyzed and, hence, they are not questioned. What is nevertheless undeniable is that the fact that most papers are written in Spanish substantially limits the dissemination of the ideas put forth in those papers among the global scientific community.

[10] This includes all six areas into which Psychology is administratively divided in Spain: Personality–Assessment–Clinical Psychology, Biological Psychology, Social Psychology, Methodology, Basic Psychology, and Developmental–Educational Psychology.

Differential Functioning of the h Index

We will use this term, intentionally borrowed from Item Response Theory, as an alternative to bias. Differential functioning refers to the fact that the *h* index has a different range and distribution across groups defined according to variables that cannot be reasonably regarded as relevant. Our analysis in this paper has sought and confirmed differential functioning across groups defined by obvious variables such as gender and age. If the current wave of exquisite concern with equity and fairness were to be carried over to other not-so-obvious but equally irrelevant variables, one could think of some that might also lower *h* indices. For instance, it is well known that *h* favors authors who generally publish with multiple co-authors. Also, scientists who are unlucky enough to suffer from a chronic disease or a physical handicap (some of whom are to be found among members of our target population) are also likely to have, as a group, lower *h* indices than their healthier colleagues. A fair, perceptive, and non-discriminative use of the *h* index should consider these and other variables along with the obvious ones of gender and age.

Conclusion

This paper has identified and demonstrated a number of weaknesses of the *h* index as a tool for the evaluation of research performance in the field of Methodology of the Behavioral Sciences, an area where *h* indices tend to be low but also where idiosyncratic publication and citation practices appear to contaminate *h*. From these results, we find little justification for the use of *h* as a criterion for grant awarding or for decisions on academic placement or tenure in fields where *h* shows these characteristics. Indeed, the total number of (qualifying) papers published by an individual appears to be a more direct indicator of scientific productivity which, at the same time, is highly linearly related to *h* (see Fig. 9). Admittedly, this measure of productivity does not carry any information about the impact that this production has had on the global scientific community, but the total number of citations conveys this information and it is also highly related to *h* (see Fig. 7d). At least, these two measures have a larger range than *h* and, hence, allow for finer distinctions than can be made with *h*, which is much more subject to spurious influences than any of these two other direct measures.

Our WoS search has come across difficulties that have been pointed out by other authors facing the same task. In the end, the true value of *h* can only be obtained by each incumbent author (who knows his/her list of publications), but then only if there is a perfect source of citation information. Given the multiplicity of decisions to be made as to which papers qualify for inclusion in the list of publications whose citation record is to be determined and as to which citing papers qualify for the final count, a thorough protocol for the computation of *h* is needed, and one that is more detailed than that put forth by Rodríguez Navarro and Imperial Ródenas (2007).

In any case, the simplicity with which bibliometricians, scientometricians, and scientists in general approach the measurement of an individual's research performance does not do justice to the complexity and multidimensionality of scientific activity, and contrasts with the richness of parameters and the depth of quantitative analysis with which the performance of tennis, basketball, or baseball players is described. Unlike patrons of these and other sports events (who would never regard an NBA player's performance as measured only by, say, his average number of offensive rebounds per game), theorists and practitioners in scientometrics seem to be content with summarizing the complexity of someone's scientific career by a single number.

References

Batista, P. D. Campiteli, M. G., Kinouchi, O., & Martinez, A. S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, *68*, 179–189.

Berry, K. J. & Mielke, P. W., Jr. (1986). R by C chi-square analyses with small expected cell frequencies. *Educational and Psychological Measurement*, *46*, 169–173.

Berry, K. J. & Mielke, P. W., Jr. (1988). Monte Carlo comparisons of the asymptotic chi-square and likelihood-ratio tests with the nonasymptotic chi-square test for sparse $r \times c$ tables. *Psychological Bulletin*, *103*, 256–264.

Bollen, J., Rodriguez, M. A., & van de Sompel, H. (2006). Journal status. *Scientometrics*, *69*, 669–687.

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

Brookfield, J. (2003). The system rewards a dishonest approach. *Nature*, *423*, 480.

Brown, N. L. (1999). On the trail of the prolific Dr Path. *Nature*, *398*, 555.

Colquhoun, D. (2003). Challenging the tyranny of impact factors. *Nature*, *423*, 479.

Falmagne, J.-C. (2005). Mathematical Psychology – A perspective. *Journal of Mathematical Psychology*, *49*, 436–439.

García-Pérez, M. A. (2000). Assessors' odd listings don't inspire confidence. *Nature*, *406*, 343.

García-Pérez, M. A. (2001). The decade 1989–1998 in Spanish psychology: An analysis of research in statistics, methodology, and psychometric theory. *Spanish Journal of Psychology*, *4*, 111–122.

García-Pérez, M. A. (2008, July). The *h* index as a measure of research accomplishments in the field of Methodology of the Behavioral Sciences. Paper presented at the III European Congress of Methodology. Oviedo, Spain.

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press

García-Pérez, M. A. (2009). A multidimensional extension to Hirsch's *h* index. Scientometrics, in press.

García-Pérez, M. A. & Núñez-Antón, V. (2009). Accuracy of power-divergence statistics for testing independence and homogeneity in two-way contingency tables. *Communications in Statistics – Simulation and Computation*, *38*, 503–512.

Hayes, A. F. & Cai, L. (2007). Further evaluating the conditional decision rule for comparing two independent means. *British Journal of Mathematical and Statistical Psychology*, *60*, 217–244.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the U.S.A.*, *102*, 16569–16572.

Hirsch, J. E. (2007). Does the *h* index have predictive power? *Proceedings of the National Academy of Sciences of the U.S.A.*, *104*, 19193–19198.

Ioannidis, J. P. A. (2008). Measuring co-authorship and networking-adjusted scientific impact. *PLoS ONE*, *3*(7), e2778. doi:10.1371/journal.pone.0002778.

Kelly, C. D. & Jennions, M. D. (2006). The *h* index and career assessment by numbers. *Trends in Ecology and Evolution*, *21*, 167–170.

Kelly, C. D. & Jennions, M. D. (2007). H-index: Age and sex make it unreliable. *Nature*, *449*, 403.

Kotiaho, J. S. (1999). Papers vanish in mis-citation black hole. *Nature*, *398*, 19.

Kotiaho, J. S., Tomkins, J. L., & Simmons, L. W. (1999). Unfamiliar citations breed mistakes. *Nature*, *400*, 307.

Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2007). Measures for measures. *Nature*, *444*, 1003–1004.

Moed, H. F. (2002). The impact-factors debate: The ISI's uses and limits. *Nature*, *415*, 731–732.

Musi-Lechuga, B., Olivas-Ávila, J. A., Portillo-Reyes, V., & Villalobos-Galvis, F. (2005). Producción de los profesores funcionarios de Psicología en España en artículos de revistas con factor de impacto en la Web of Science. *Psicothema*, *17*, 539–548.

Nature (2005a). Not-so-deep impact [Editorial]. *Nature*, *435*, 1003–1004.

Nature (2005b). Ratings games [Editorial]. *Nature*, *436*, 889–890.

Price, N. C. (1998). What's in a name (or a number or a date)? *Nature*, *395*, 538.

van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, *67*, 491–502.

Rodríguez Navarro, A. & Imperial Ródenas, J. (2007). *Índice h. Guía Para la Evaluación de la Investigación Española en Ciencia y Tecnología Utilizando el Índice h*. Madrid: Consejería de Educación de la Comunidad de Madrid. Electronic document available at http://www.madrimasd.org/informacionidi/biblioteca/publicacion/doc/33_indiceh.zip

Salgado, J. F. & Páez, D. (2007). La productividad científica y el índice *h* de Hirchs [*sic*] de la psicología social española: Convergencia entre indicadores de productividad y comparación con otras áreas. *Psicothema*, *19*, 179–189.

Schreiber, M. (2007). A case study of the Hirsch index for 26 non-prominent physicists. *Annalen der Physik*, *16*, 640–652.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, *314*, 498–502.

Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized Hirsch *h*-index for disclosing latent facts in citation networks. *Scientometrics*, *72*, 253–280.

Taber, D. F. (2005). Quantifying publication impact. *Science*, *309*, 2166.

Todd, P. A. & Ladle, R. J. (2008). Citations: Poor practices by authors reduce their value. *Nature*, *451*, 244.

Vinkler, P. (2007). Eminence of scientists in the light of the *h*-index and other scientometric indicators. *Journal of Information Science*, *33*, 481–491.

Waheed, A. A. (2003). Citation rate unrelated to journals' impact factors. *Nature*, *426*, 495.

Wendl, M. C. (2007). H-index: However ranked, citations need context. *Nature*, *449*, 403.

Yellott, J. I., Jr. (1969). Probability learning with noncontingent success. *Journal of Mathematical Psychology*, *6*, 541–575.

Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, *57*, 173–181.

Received October 2, 2008
Revision received November 24, 2008
Accepted December 4, 2008

APPENDIX

To test for homogeneity of the distributions of h among eligible full and associate professors (Figs. 1a and 1b), the data were arranged in a 2×9 contingency table where each row contains the frequency distribution of h indices in one of the groups of professors, with the first eight columns including h values from 0 to 7 and the last column including h values of 8 or higher. The value of Pearson's X^2 statistic was 15.47 with 8 degrees of freedom, but there were 9 (out of 18) expected frequencies below 5, four of which were also below 1. In these conditions, Pearson's statistic does not follow its asymptotic χ^2 distribution. Berry and Mielke (1988) derived a non-asymptotic Pearson Type-III distribution that presumably holds in cases like this, and García-Pérez and Núñez-Antón (2009) have further demonstrated that this is indeed the case regardless of the number of small expected frequencies and also regardless of how small they are, provided that table density (i.e., the ratio of sample size to number of cells in the table) exceeds unity. In our case, table density is 147/18 = 8.17. Then, the p-value of our sample Pearson's X^2 statistic, computed with the software described by Berry and Mielke (1986) is 0.057, so that the distribution of h among full professors is not significantly different from its distribution among associate professors

https://doi.org/10.1017/S1138741600002201 Published online by Cambridge University Press