WILEY InterScience® DISCOVER SOMETHING GREAT

Are There Better Indices for Evaluation Purposes than the *h* Index? A Comparison of Nine Different Variants of the *h* Index Using Data from Biomedicine

Lutz Bornmann
ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr. 24, CH-8092 Zurich. E-mail: bornmann@gess.ethz.ch

Rüdiger Mutz
ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr. 24, CH-8092 Zurich. E-mail: mutz@gess.ethz.ch

Hans-Dieter Daniel
ETH Zurich, Professorship for Social Psychology and Research on Higher Education, Zähringerstr. 24, CH-8092 Zurich. University of Zurich, Evaluation Office, Mühlegasse 21, CH-8001 Zurich. E-mail: daniel@evaluation.uzh.ch

In this study, we examined empirical results on the *h* index and its most important variants in order to determine whether the variants developed are associated with an incremental contribution for evaluation purposes. The results of a factor analysis using bibliographic data on postdoctoral researchers in biomedicine indicate that regarding the *h* index and its variants, we are dealing with two types of indices that load on one factor each. One type describes the most productive core of a scientist's output and gives the number of papers in that core. The other type of indices describes the impact of the papers in the core. Because an index for evaluative purposes is a useful yardstick for comparison among scientists if the index corresponds strongly with peer assessments, we calculated a logistic regression analysis with the two factors resulting from the factor analysis as independent variables and peer assessment of the postdoctoral researchers as the dependent variable. The results of the regression analysis show that peer assessments can be predicted better using the factor 'impact of the productive core' than using the factor 'quantity of the productive core.'

Introduction

Jorge Hirsch (2005) proposed the *h* index as a criterion to quantify the scientific output of a single researcher. The *h* index was put forward as a better alternative to other citation-based metrics that could be used to measure research achievement (for example, total number of citations or citations per paper). Hirsch's (2005) *h* index depends on both the number of a scientist's publications and the impact of the papers on the scientist's peers. "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 fewer than $\leq h$ citations each" (Hirsch, 2005, p. 16569). The proposed new measure of research performance was quickly taken up by *Nature* (Ball, 2005) and *Science* (Anon, 2005). The idea of ranking scientists by a single number, as well as the alleged advantages that the *h* index has over other citation-based indicators, attracted the attention of science news editors and researchers working in the area of information sciences and evaluative bibliometrics. In a review of the research literature on the *h* index, Bornmann and Daniel (2007b) found 30 papers published one year after Hirsch's 2005 article. Furthermore, "an entire issue of the journal *Scientometrics* was recently devoted to the *h*-index, and the measure is now automatically calculated in the 'citation report' function of *Web of Science*" (Meho, 2007, p. 36). In the face of the popularity of the *h* index, it has been suggested that as a measurement of the impact of individual scientists, it is equivalent to the Journal Impact Factor for journals (provided by Thomson Scientific, Philadelphia, PA, USA; Gracza & Somoskovi, 2007).

The *h* index is seen to have the advantage that it gives a robust estimate of the *broad* impact of a scientist's cumulative research contributions (Hirsch, 2005). It combines

Received September 26, 2007; revised November 14, 2007; accepted November 19, 2007

 • Published online 7 February 2008 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/asi.20806

number of publications and citation counts in a “balanced way” (Braun, Glänzel, & Schubert, 2005), being insensitive to a set of lowly cited (noncited) papers or to one or several highly cited papers. That means that the *h* index favors “those authors who produce a series of influential papers rather than those authors who either produce many papers that are soon forgotten or produce a few that are uncharacteristically influential” (Kelly & Jennions, 2006, p. 167). The *h* index can be used not only to determine the past productivity of a scientist, but it is also—as Hirsch (2007) ascertains in his most recent published study—better able than other bibliometric indicators to predict a scientist’s future productivity (Ball, 2007). For this reason, Hirsch (2007) finds the *h* index “a useful indicator of scientific quality that can be profitably used (together with other criteria) to assist in academic appointment processes and to allocate research resources” (p. 19198). According to Glänzel (2006), another strength of the *h* index is that for the assessment of small paper sets (as was the case for the postdoctoral researchers examined in the present study) it is particularly well suited when “other, traditional bibliometric indicators often fail or at least were [*sic!*] their application proved usually problematic” (p. 320).

Besides the advantages, a number of disadvantages of the *h* index have been named in recent months (for an overview, see Bornmann & Daniel, 2007b; Jin, Liang, Rousseau, & Egghe, 2007; Liu & Rousseau, 2007). It has been pointed out that the *h* index is only weakly sensitive to the number of citations received by single publications. According to Egghe (2006a), “such an index should be sensitive to the level of the highly cited papers. As the *h* index is defined now, once an article belongs to the *h*-defining class, it is totally unimportant whether or not these papers continue to be cited and, if cited, it is unimportant whether these papers receive 10, 100, or 1,000 more citations” (p. 14). Also, as the *h* index is highly dependent upon a scientist’s number of years of active research, strictly speaking we should only compare the *h* values of scientists that have been active researchers for a similar number of years. Finally, a scientist’s *h* value can only rise and cannot decline. Scientists that never publish another paper, or that are inactive researchers, maintain (at worst) the same value on the *h* index. For these reasons, Jin et al. (2007) find it “no surprise that colleagues proposed . . . some new ‘Hirsch-type’ indices trying to overcome some of the disadvantages” (p. 856).

We examined the *h* index and the most important Hirsch-type indices, or *h* index variants, that have been proposed and discussed in the literature: the *m* quotient (Hirsch, 2005), *g* index (Egghe, 2006b), *h(2)* index (Kosmulski, 2006), *a* index (Jin, 2006), *r* index (Jin et al., 2007), *ar* index (Jin et al., 2007), and h_w index (Egghe & Rousseau, in press). We also included in our analysis the *m* index, a variant that we propose of the *a* index. The definitions of all of these variants are provided below. The aim of the analysis here is to determine empirically the extent to which the development of the variants of the *h* index does in fact result in an incremental contribution. Although the proposed variants may be conceptualized differently than the *h* index theoretically or mathematically, in their empirical application they may be highly correlated with the *h* index and with each other, and it may be that they are but simple linear transformations of each other (Occam’s razor principle). If the present study were to reveal high intercorrelations, the indices could be called redundant in empirical application.

Methods

Statistical Analysis

What psychometricians do first with a battery of diverse metrics that are purported to measure something similar in a particular way is to look at their intercorrelational structure using factor analysis (Harnad, 2007). Factor analysis is a statistical method “to reduce the dimensionality of the data space in order to discover, visualize, and interpret dependencies among sets of variables” (Timm, 2002, p. 445). The factor analysis provides information on the dimensionality of the structure of the dependencies in a data set (1-dimensional, 2-dimensional, . . .). With regard to the different indices for measuring the scientific output of researchers (the *h* index and its variants), we therefore examined in this study whether the indices are based on a single factor (such as a research performance dimension) or whether there are several factors involved (such as, in addition to a research performance dimension, a time-related dimension, reflecting number of years of active research). The variants of the *h* index would provide incremental contribution only if the factor analysis showed not only one dimension but several dimensions.

In addition to examining the structure of the dependencies, by examining in the present study the incremental contribution of the individual indices, we are also looking into the question recently raised by Burrell (2007a): “From a practical point of view, which index is to be preferred?” (p. 168). Seeking an answer to this question, we examined which of the indices are useful yardsticks to compare different scientists for evaluative purposes. Using multiple logistic regression analysis (see Hosmer & Lemeshow, 2000), we found out what indices are significant predictors of peer assessments of different scientists.

In the present study, we looked at the most important variants of the *h* index that have been discussed in greater detail in the literature: the *m* quotient, *g* index, *h(2)* index, *a* index, *r* index, *ar* index, and h_w index (see Table 1). Some alternatives to the *h* index, which are not relevant for the data set examined in this study, were not considered: As all of the scientists for whom indices were calculated in the present study conduct biomedical research, there is no need to compare index values across scientific disciplines. For this reason, we did not use the standardizations of the *h* index for comparing scientists that work in different scientific fields developed by Batista, Campiteli, and Kinouchi (2006), Iglesias and Pecharroman (2007a, 2007b), Imperial and Rodríguez-Navarro (2007), and Levitt and Thelwall (2007). And as the scientists in this study are without exception

DOI: 10.1002/asi

TABLE 1. Definitions of the *h* index and its variants.

Index	Definition
h index	"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 fewer than $\leq h$ citations each" (Hirsch, 2005, p. 16569)
m quotient	$\frac{h}{y}$ where $h = h$ index, $y =$ number of years since publishing the first paper
g index	"The highest number *g* of papers that together received g^2 or more citations" (Egghe, 2006b)
h(2) index	"A scientist's *h*(2) index is defined as the highest natural number such that his *h*(2) most-cited papers received each at least $[h(2)]^2$ citations" (Kosmulski, 2006, p. 4)
a index	$\frac{1}{h}\sum_{j=1}^{h} cit_j$ where $h = h$ index, $cit =$ citation counts
m index	The median number of citations received by papers in the Hirsch core (this is the papers ranking smaller than or equal to *h*)
r index	$\sqrt{\sum_{j=1}^{h} cit_j}$ where $h = h$ index, $cit =$ citation counts
ar index	$\sqrt{\sum_{j=1}^{h} \frac{cit_j}{a_j}}$ where $h = h$ index, $cit =$ citation counts, $a =$ number of years since publishing
h_w index	$\sqrt{\sum_{j=1}^{r_0} cit_j}$ where $cit =$ citation counts, $r_o =$ the largest row index *j* such that $r_w(j) \leq cit_j$

active researchers, we did not compute the contemporary *h* index as proposed by Sidiropoulos, Katsaros, and Manolopoulos (2007). With the contemporary *h* index, the authors defined a generalization of the *h* index in order to account for the fact that "senior scientists, who keep contributing nowadays, or brilliant young scientists, who are expected to contribute a large number of significant works in the near future but now they have only a small number of important articles due to the time constraint, are not distinguished by the original *h*-index" (p. 257, see here also Harzing, 2007).

The Data Set for the Investigation of the Indices

We investigated committee peer review for awarding long-term fellowships to postdoctoral researchers as practiced by the Boehringer Ingelheim Fonds (B.I.F.; www.bifonds.de)—an international foundation for the promotion of basic research in biomedicine (Bornmann & Daniel, 2005b, 2006). According to Fröhlich (2001), managing director of the B.I.F., applicants that demonstrate excellence in scientific work are selected for the fellowships by the B.I.F. Board of Trustees (seven internationally renowned scientists); otherwise the applicants are rejected. Our evaluation study involved 414 postdoctoral applicants (64 approved and 350 rejected) from the years 1990 to 1995, with a total of 1,586 papers that they published previous to applying for the fellowship (publication window: 1986–1994). The papers received a total of 60,882 citations (according to the Science Citation Index provided by Thomson Scientific; citation window: from year of publication to the end of 2001).

Definitions of the h index Variants

m quotient

According to a stochastic model for an author's production/citation process, Burrell (2007b) conjectures that the *h* index is approximately proportional to career length. One way to compare scientists with different lengths of scientific careers is to divide the *h* index by number of years of research activity. For this reason, Hirsch (2005) already proposed dividing the *h* index by number of years since a scientist's first publication and called this quotient *m*.

g index

Holding that "a measure which should indicate the overall quality of a scientist . . . should deal with the performance of the top articles," Egghe (2006b. p.1) proposed the *g* index as a modification of the *h* index. For the calculation of the *g* index, the same ranking of a publication set—papers in decreasing order of the number of citations received—is used as for the *h* index (Jin et al., 2007). Egghe (2006b) defines the *g* index "as the highest number *g* of papers that together received g^2 or more citations. From this definition it is already clear that $g \geq h$." In contrast to the *h* index, the *g* index gives more weight to highly cited papers. The aim is to avoid a disadvantage of the *h* index that "once a paper belongs to the top *h* papers, its subsequent citations no longer 'count' " (Harzing, 2007). Egghe (2006c) calculated *h* index values and *g* index values for (still active) Derek de Solla Price Medallists and found that "the ranked *g*-index column resembles more the overall feeling of 'visibility' or 'life time achievement' than does the ranked *h*-index column" (p. 144).

h(2) index

Like the *g* index, calculation of the *h(2)* index also gives more weight to highly cited articles: "A scientist's *h*(2) index is defined as the highest natural number such that his *h*(2) most-cited papers received each at least $[h(2)]^2$ citations" (Kosmulski, 2006, p. 4). An *h(2)* index of 20, for example, means that a scientist has published at least 20 papers, of which each has been cited at least 400 times. Obviously, for any scientist, the *h(2)* index is always lower than the *h* index. According to Jin

DOI: 10.1002/asi

et al. (2007), the main advantage of the *h(2)* index is "that it reduces the precision problem" (p. 856). That means that when computing the *h(2)* index using a publication set put together for a scientist using Web of Science data (Thomson Scientific), less work is needed to check the accuracy of the publications data, especially with regard to homographs—that is, to distinguish between scientists that have the same last name and first initial (see Meho, 2007)—than is needed when calculating the *h* index (see here also Liu & Rousseau, 2007; Schreiber, 2007). As only few papers in the set are sufficiently highly cited in order to fulfill the criterion of $[h(2)]^2$ citations, there are also fewer papers to check.

a index

According to Burrell (2007c) "the *h*-index seeks to identify the most productive core of an author's output in terms of most received citations" (p. 170). For this core, consisting of the first *h* papers, Rousseau (2006) introduced the term *Hirsch core*. "The Hirsch core can be considered as a group of high-performance publications, with respect to the scientist's career" (Jin et al., 2007, p. 855). The *a* index (as well as the *m* index, *r index*, and *ar index* described below) includes in the calculation only papers that are in the Hirsch core; it is defined as the average number of citations of papers in the Hirsch core. The proposal to use this average number of citations as a variant of the *h* index was made by Jin, the main editor of *Science Focus* (Jin, 2006). Rousseau (2006) referred to this index later as the *a* index.

m index

As the distribution of citation counts is usually skewed, the median and not the arithmetic average should be used as the measure of central tendency. Therefore, as a variation of the *a* index, we propose the *m* index—the median number of citations received by papers in the Hirsch core.

r index

Jin et al. (2007) observed critically that with the *a* index, "the better scientist is 'punished' for having a higher *h*-index, as the *A*-index involves a division by *h*" (p. 857). Therefore, instead of dividing by *h*, the authors suggest taking the square root of the sum of citations in the Hirsch core to calculate the index. Jin et al. (2007) refer to this new index as the *r* index, as it is calculated using a square root. As the *r* index—similar to the *a* index—measures the citation intensity in the Hirsch core, the index can be very sensitive to just a very few papers receiving extremely high citation counts.

ar index

The *ar* index is an adaptation of the *r* index. It takes into account not only the citation intensity in the Hirsch core but also makes use of the age of the publications in the core (Jin et al., 2007). This is an index that not only can increase but also decrease over time. For a good research evaluation indicator, Jin et al. (2007) see it as a necessary condition that the index has sensitivity to performance changes. For this reason, Jin (2007) proposes the *ar* index, "defined as the square root of the sum of the average number of citations per year of articles included in the *h*-core." To illustrate the necessity of a decreasing index in concrete application, Jin et al. (2007) calculated the *h* index, *r* index, and the *ar* index for the articles written by B. C. Brookes (Brookes, who was the Derek de Solla Price Medallist in 1989, died in 1991): "Brookes' *h*-index over the whole period (2002–2007) stays fixed at $h = 12$ (hence here $h > AR$). Between 2002 and 2007 his *R*-index increased by 5% while the *AR*-index decreased by about 5%" (Jin et al., 2007, p. 859).

h_w *index*

Similar to the *ar* index, the h_w index (an *h* index weighted by citation impact) developed by Egghe and Rousseau (in press) is sensitive to performance changes. The h_w index is defined as:

$$\sqrt{\sum_{j=1}^{r_0} cit_j}$$

where r_o is the largest row index *j* such that $r_w(j) \leq cit_j$.

Results

All of the nine indices that were included in the present study were calculated for the B.I.F. applicants following the definitions in Table 1 and using the statistical software package SPSS (SPSS for Windows, 2006). The syntax for calculation of the individual indices was developed by the first author of this paper and is available upon request. The factor analysis and the logistic regression analysis, the results of which are reported below, were calculated using MPLUS 4.2 (Muthén & Muthén, 1998–2006) and SAS 9.1.3 (SAS Institute Inc., 2007).

Descriptive Statistics

Table 2 shows the statistical values *min*, *max*, *m*, *sd*, and *mdn* of the nine indices for the B.I.F. applicants in overview. As the table shows, the *h* index values for the applicants lie between 0 and 13. The *m* and the *sd* of the *h* index values are 2.89 and 2.23, and the *mdn* has a value of 2. In agreement with Liu and Rousseau (2007), the mean (*m*) *h* index value is larger than the mean *h*(2) index value, and it is smaller than the mean *g* index value (see Table 2). Although a number of B.I.F. applicants published only a few articles, some of the articles were highly cited, so that the mean *a* index and *m*

TABLE 2. Minimum (*min*), maximum (*max*), mean (*m*), standard deviation (*sd*), and median (*mdn*) of nine indices calculated for the B.I.F. fellowship applicants.

Index	*n*	*min*	*max*	*m*	*sd*	*mdn*
h index	414	0	13.00	2.89	2.23	2.00
m quotient	414	0	1.20	0.26	0.19	0.22
g index	414	0	21.00	3.52	2.97	3.00
h(2) index	414	0	7.00	2.23	1.42	2.00
a index	414	0	434.00	41.51	52.38	23.52
m index	414	0	434.00	36.96	49.98	21.00
r index	414	0	42.10	9.52	7.43	8.09
ar index	414	0	15.64	3.09	2.41	2.54
h_w index	414	0	38.46	8.66	6.67	7.55

DOI: 10.1002/asi

index values depart very strongly from the mean *h* index value, as the table shows.

Scale Transformation of the Indices

As the median (*mdn*) of the individual indices in Table 2 in part deviates very strongly from the mean (*m;* the *a* index, for example) and as the Shapiro-Wilks normality test is statistically significant for all of the indices, the indices are not symmetrically distributed and not normally distributed. According to Egghe (2005a; 2005b) it can be assumed, in addition, that the relationship between any two indices (y, x) is nonlinear and can be described using a power function: $y = \mathrm{f}(x) = C * x^{-a}$, where C and a are constants. Logarithmizing the equation results in a simple linear function: $y = \log_e(\mathrm{f}(x)) = \log_e(C) - a * \log_e(x)$. Therefore, all of the index values of the B.I.F applicants—following the equation—were logarithmized ($\log_e(x + 1)$). This logarithmic transformation has the additional effect that the distribution of data more likely approximates a normal distribution.

Figure 1 shows that after the transformation the data seem to be approximately normally distributed; still, considering the significant Shapiro-Wilks normality tests, the assumption of normally distributed data also after the logarithmic transformation is not fully met. Therefore, a special variant of maximum likelihood (ML) factor analysis is used, which produces parameter estimates with mean-adjusted χ^2 test values (Satorra-Bentler correction) that are robust to non-normality (Curran, West, & Finch, 1996; Muthén & Muthén, 1998–2006).

Results of an Exploratory Factor Analysis

Using an exploratory factor analysis, we tried to find basic dimensions (or factors) that indicate how the nine indices (the *h* index and its variants) calculated for the B.I.F. applicants cluster (Kline, 1998; Stevens, 1996). By the use of maximum likelihood principal component analysis (MLPCA), our set of correlated variables (the indices) was transformed into a set of uncorrelated latent variables (dimensions or factors). Kaiser's eigenvalue criterion (eigenvalue > 1.0) and explained variance greater than 95% were used to determine the number of factors. The mean-adjusted χ^2 test value decreased from 2473.0 (one-factor solution) to 904.5 (two-factor solution). The result of the analysis was a factor loading matrix that indicates the degree of correlation between the indices and the factors. We used a rotated VARIMAX-transformation to make the matrix more easily interpretable.

The results of the factor analysis show that two factors with eigenvalues larger than 1 explain more than 95% of the total variance in the matrix of the correlations between the indices. The factor loading matrix for the two factors and the nine indices is shown in Table 3. The categorization of the indices among the factors by using factor loadings greater than 0.6 (these loadings are shown in bold face in the table) revealed that the indices can be categorized in terms of their relation to the most productive core of the output of an applicant. Whereas indices that load on the first factor (that is, the *h* index, *m* quotient, *g* index, and *h*(2)index) indicate the *number* of papers in a defined most productive core of the output of an applicant, indices that load on the second factor (that is, the *a* index, *m* index, *r* index, *ar* index, and h_w index) quantify the *impact* of the papers in that core. Based on the assignment of the indices to the factors, we refer to factor 1 as 'quantity of the productive core' and to factor 2 as 'impact of the productive core' (see Table 3). Factor 1 explains 47.3% and factor 2 explains 48.4% of the variance in the matrix of correlations. Hence, the factors 'quantity of the productive core' and 'impact of the productive core' explain a similar percentage of the variance in the data. As Table 3 shows, of all of the indices, the *h* index and the *g* index load the highest on factor 1 with coefficients of 0.91 and 0.94, and the *m* index and the *a* index load the highest on factor 2 with a coefficient in each case of 0.96.

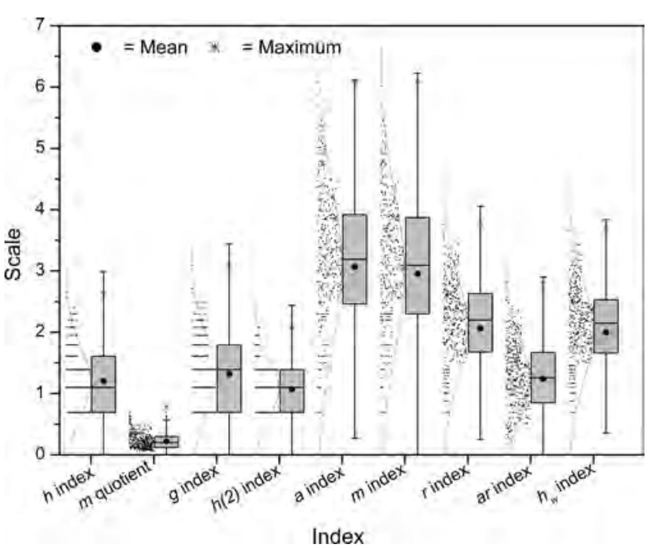

FIG. 1. Box plots of the nine logarithmic indices calculated for the B.I.F. applicants with the distribution of data and the Gaussian bell curve.

TABLE 3. Varimax rotated loading matrix of a ML factor analysis for nine log transformed indices ($n = 414$).

Index	Factor 1 ('quantity of the productive core')	Factor 2 ('impact of the productive core')	Communality
h index	**0.94**	0.34	1.00
m quotient	**0.88**	0.27	0.85
g index	**0.91**	0.32	0.93
h(2) index	**0.83**	0.50	0.94
a index	0.28	**0.96**	1.00
m index	0.22	**0.96**	0.97
r index	0.58	**0.82**	1.00
ar index	0.56	**0.80**	0.95
h_w index	0.56	**0.82**	0.99
Eigenvalue	4.26	4.36	
Explained variance [%]	47.3	48.4	

Note. Values in bold face denote factor loadings greater than 0.60.

DOI: 10.1002/asi

Results of a Logistic Regression Analysis

An index for measuring research output is a useful yardstick to compare different scientists, if the index is strongly related to peer assessments (see here Cole, 1989). The fellowship selection decisions of the B.I.F. Board of Trustees (approval or rejection) for the individual B.I.F. applicants, for whom we calculated the nine indices, were available to us for investigation. To determine the relationship between the indices and the Trustees' decisions on the applicants, we calculated a multiple logistic regression analysis (Hosmer & Lemeshow, 2000). A multiple logistic regression model is appropriate for the analysis of dichotomous responses. In the case of the B.I.F. applicants, the binary response is coded 0 for approval and 1 for rejection of a fellowship application. As the high intercorrelation of the individual indices (multicollinearity) leads to problems in the estimation of the regression coefficients, we did not include the indices themselves in the analysis as predictors but instead the two factors that resulted from the factor analysis. This provides information on the extent to which the indices that correlate highly with the 'impact of the productive core' and/or the indices that correlate highly with the 'quantity of the productive core' are related to the B.I.F. Trustees' assessments.

Table 4 shows the results of the logistic regression analysis, in which the two factors ('quantity of the productive core' and 'impact of the productive core') were included as independent variables and the selection decision of the Board of Trustees was included as the dependent variable. The overall result of the analysis is significant ($\text{Wald}_{\text{total}}\text{-}\chi^2(2) = 23.97$, $p < 0.05$, $R^2 = 0.13$). Both factors included predict the Trustees' selection decision statistically significantly ('quantity of the productive core': Wald-$\chi^2(1) = 9.71$, $p < 0.05$; 'impact of the productive core': Wald-$\chi^2(1) = 22.15$, $p < 0.05$). As the factor 'impact of the productive core' ($R^2 = 0.09$) can explain nearly twice the variance in the data than the factor 'quantity of the productive core' ($R^2 = 0.04$), the factor 'impact of the productive core' (and thus the indices that load on this factor) can better predict peer assessments than the factor 'quantity of the productive core.' We can illustrate the effect of this difference between the factors in practical application taking the example of a comparison of the *h* index and the *m* index with regard to the selection decisions on the B.I.F. applicants. Whereas the *h* index, as compared to the other indices, loads the highest on the factor 'quantity of the productive core,' the *m* index, as compared to the other indices, loads the highest (in conjunction with the *a* index) on the factor 'impact of the productive core' (see the results in Table 3). Both of the variables represent the marker variables for one of the two factors.

Table 5 (top) shows the relations between the B.I.F. applicants' mean *h* index values (*m, mdn*) and the decisions (approval or rejection) of the Board of Trustees for the years 1990 to 1995. For every year, *m* and *mdn* of approved applicants are consistently higher than those of rejected applicants. To determine the strength of the association between the *h* index values and the Board's decisions, we employed a measure of association, Cramer's V (see Kline, 2004). Values of Cramer's V between 0.32 and 0.61 in Table 5 (top) indicate medium to strong effect sizes for the relation between *h* index values and decisions made by the Board of Trustees (see Cohen, 1988). These results are reported in Bornmann and Daniel (2005a, 2007a). Table 5 (bottom) shows the relations between the applicants' mean *m* index values (*m, mdn*) and the decisions of the Board of Trustees for the individual years. Here again, *m* and *mdn* values of approved applicants are consistently higher than those of rejected applicants.

TABLE 4. Summary of a multiple logistic regression analysis of the decision of the B.I.F. Board of Trustees (approval = 0 or rejection = 1) on the two factors that resulted from the factor analysis in TABLE 3 ($n = 414$).

Parameter	*df*	Estimate	Wald-χ^2	R^2 (new-scaled)
Total	2		23.97*	0.13
Intercept	1	1.97	130.96*	
Factor 1 ('quantity of the productive core')	1	−0.47	9.71*	
Factor 2 ('impact of the productive core')	1	−0.86	22.15*	

* $p < 0.05$

TABLE 5. Arithmetic mean (*m*), standard deviation (*sd*), and median (*mdn*) of *h* index and *m* index for approved and rejected B.I.F. applicants by year of Board of Trustees' meeting. The value of Cramer's V indicates the effect size for the relation between an index and the Trustees' decision in a particular application year.

	Year of Board of Trustees' meeting					
	1990	1991	1992	1993	1994	1995
***h* index**						
Approved						
m	5.15	3.90	2.92	4.14	2.83	4.33
sd	3.13	3.35	2.29	2.85	1.27	2.06
mdn	4.00	3.00	3.00	3.00	3.00	5.00
n	13	10	13	7	12	9
Rejected						
m	2.71	2.94	2.70	2.40	2.46	2.99
sd	2.58	2.12	2.17	1.69	2.11	2.05
mdn	2.00	2.00	2.00	2.00	2.00	3.00
n	52	36	57	60	52	93
Cramer's V	0.61	0.52	0.41	0.52	0.32	0.35
***m* index**						
Approved						
m	94.65	48.55	44.35	54.64	73.58	85.17
sd	88.27	47.79	42.75	27.41	67.60	127.58
mdn	88.00	31.50	26.00	45.00	56.00	21.50
n	13	10	13	7	12	9
Rejected						
m	23.71	31.60	29.86	22.42	45.36	34.41
sd	30.16	29.53	39.24	20.73	56.51	51.59
mdn	16.75	25.00	16.00	14.25	23.50	20.00
n	52	36	57	60	52	93
Cramer's V	0.97	0.83	0.87	0.90	0.95	0.85

Note. Publication window: 1986–1994; citation window: from year of publication to the end of 2001.

DOI: 10.1002/asi

However, the values of Cramer's V (between 0.83 and 0.97) in all years indicate a clearly closer relationship between index and selection decision than was the case with the *h* index. Due to the low applicant numbers, the values of Cramer's V in Table 5 have to be interpreted with great care. In order to test whether for the individual application years the stronger association between the Board of Trustees' decisions and the *m* index values (than between Board's decisions and the *h* index values) is also found when a different measure of association is used, we calculated Eta values. The results again showed higher values for the *m* index than for the *h* index throughout. All in all, these findings mean that the *m* index allows better discrimination between approved and rejected applicants than the *h* index.

Discussion

Due to the disadvantages of the *h* index that have been named since Hirsch's first publication of the index in 2005 (Hirsch, 2005), a number of variants that are intended to compensate for the weaknesses have been proposed: *m* quotient, *g* index, *h(2)* index, *a* index, *m* index (a variation of the *a* index proposed by the present authors), *r* index, *ar* index, and *hw* index. Although numerous studies on the indices have already been published (see an overview in Bornmann & Daniel, 2007b), we found only three studies that tested the relationship between the indices using statistical methods: Kosmulski (2006) examined the relationship between the *h* index and the *h(2)* index of 19 full professors affiliated at a department of chemistry of a university in Poland. Kosmulsiki's results show that there is a high correlation between the two indices ($r = 0.91$). Similarly high correlations are reported also by two other studies: Jin and colleagues (Jin et al., 2007) determined the relationship between the *g* index and the *r* index for Derek de Solla Price Medallists ($r = 0.99$), and Ravichandra Rao (2007) determined the relationship between the *h* index and the *g* index for 168 authors that contributed references to a bibliography on 'optical flow estimation' ($r = 0.97$). The high correlations between the indices that were found in these studies indicate that development of the variants of the *h* index (see our starting out question in the Introduction section) has resulted in hardly any empirical incremental contribution.

In order to identify the factors that mathematically account for the variance in the correlation matrix of the nine different indices, we calculated an exploratory factor analysis in the present study. The results of the analysis indicate that with the *h* index and its variants, we can assume that there are two types of indices:

- The one type of indices (*h* index, *m* quotient, *g* index and *h(2)* index) describe the most productive core of the output of a scientist and tell us the number of papers in the core.
- The other indices (*a* index, *m* index, *r* index, *ar* index, and *hw* index) depict the impact of the papers in the core.

The two index types thus stand for very different dimensions of scientists' research output, but they can complement each other very well. Our results indicate that there is an empirical incremental contribution associated with some of the variants of the *h* index that have been proposed up to now—that is, with the variants that depict the impact of the papers in the productive core. Upon that background, Jin and colleagues' (Jin et al., 2007) recommendation that a combination of indices be used for evaluative purposes (see also Liu & Rousseau, 2007) seems very sensible. While Jin et al. (2007) propose the use of a combination of the *h* index and *r* index, or of the *h* index and the *ar* index, based on our findings we propose the use of any pair of indices as a meaningful indicator for comparing scientists, where one index relates to the number of papers in a researcher's productive core (namely, the *h* index or *g* index—that is, one of the indices with the highest loadings on this factor in the factor analysis) and the other index relates to the impact of the papers in a researcher's productive core (namely, the *a* index or *m* index—that is, one of the indices with the highest loadings on this factor in the factor analysis).

As an index for the measurement of scientific outputs is a useful yardstick for comparing scientists if the index corresponds strongly with the assessment by peers, we calculated in this study a logistic regression analysis, with both dimensions yielded by the factor analysis ('quantity of the productive core' and 'impact of the productive core') as independent variables and the B.I.F. Board of Trustees' selection decision on researchers' fellowship applications as the dependent variable. The results of the regression analysis show that the factor 'impact of the productive core' is a better predictor for the Trustees' assessments than the factor 'quantity of the productive core.' This result agrees with a statement by the managing director of the B.I.F. (Fröhlich, 2001), according to which when selecting fellowship recipients, the Board of Trustees of the B.I.F. looks for excellence in scientific performance: Excellence usually finds expression in the quality of a scientist's best publications. Even though Ball (2007) in *Nature* recently comes to the conclusion that 'the *h*-index does seem to be able to identify good scientists and it is becoming widely used informally, for example to rank applicants for research posts,' our results indicate that indices other than the *h* index are even better suited to those purposes.

However, for both the regression analysis and the factor analysis, it will be important to test whether the results of the present study can be replicated using data sets from other scientific fields. For example, the factor loadings and factor correlations found here could be defined a priori in a structural equation model and the model's fit tested with respect to other data sets from other fields (Ullman & Bentler, 2004).

Acknowledgment

The authors wish to express their gratitude to three anonymous reviewers for their helpful comments.

References

Anon. (2005). Data point. Science, 309(5738), 1181.

Ball, P. (2005). Index aims for fair ranking of scientists. Nature, 436(7053), 900.

Ball, P. (2007). Achievement index climbs the ranks. Nature, 448(7155), 737.

Batista, P.D., Campiteli, M.G., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? Scientometrics, 68(1), 179–189.

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

Bornmann, L., & Daniel, H.-D. (2005b). Selection of research fellowship recipients by committee peer review. Analysis of reliability, fairness and predictive validity of Board of Trustees' decisions. Scientometrics, 63(2), 297–320.

Bornmann, L., & Daniel, H.-D. (2006). Selecting scientific excellence through committee peer review – a citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. Scientometrics, 68(3), 427–440.

Bornmann, L., & Daniel, H.-D. (2007a). Convergent validation of peer review decisions using the *h* index: extent of and reasons for type I and type II errors. Journal of Informetrics, 1(3), 204–213.

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

Braun, T., Glänzel, W., & Schubert, A. (2005). A Hirsch-type index for journals. The Scientist, 19(22), 8.

Burrell, Q.L. (2007a). Hirsch's h-index and Egghe's g-index. In D. Torres-Salinas & H.F. Moed (Eds.), Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics (Vol. 1, pp. 162–169). Madrid, Spain: Spanish Research Council (CSIC).

Burrell, Q.L. (2007b). Hirsch's h-index: a stochastic model. Journal of Informetrics, 1(1), 16–25.

Burrell, Q.L. (2007c). On the h-index, the size of the Hirsch core and Jin's A-index. Journal of Informetrics, 1(2), 170–177.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Publishers.

Cole, S. (1989). Citations and the evaluation of individual scientists. Trends in Biochemical Sciences, 14(1), 9–13.

Curran, P.J., West, S.G., & Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Methods, 1(1), 16–29.

Egghe, L. (2005a). Power laws in the information production process: Lotkaian informetrics. Kidlington, UK: Elsevier Academic Press.

Egghe, L. (2005b). The power of power laws and an interpretation of lotkaian informetric systems as self-similar fractals. Journal of the American Society for Information Science and Technology, 56(7), 669–675.

Egghe, L. (2006a). How to improve the h-index? The Scientist, 20(3), 14.

Egghe, L. (2006b). An improvement of the h-index: the g-index. ISSI Newsletter, 2(1), 8–9.

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

Egghe, L., & Rousseau, R. (in press). An *h*-index weighted by citation impact. Information Processing & Management.

Fröhlich, H. (2001). It all depends on the individuals. Research promotion - a balanced system of control. B.I.F. Futura, 16, 69–77.

Glänzel, W. (2006). On the *h*-index – a mathematical approach to a new measure of publication activity and citation impact. Scientometrics, 67(2), 315–321.

Gracza, T., & Somoskovi, I. (2007). Impact factor and/ or Hirsch index? Orvosi Hetilap, 148(18), 849–852.

Harnad, S. (2007). Open access scientometrics and the UK research assessment exercise. In D. Torres-Salinas & H.F. Moed (Eds.), Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics (Vol. 1, pp. 27–33). Madrid, Spain: Spanish Research Council (CSIC).

Harzing, A.-W. (2007). Reflections on the h-index. Retrieved July 19, 2007, from http://www.harzing.com/pop_hindex.htm

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.

Hirsch, J.E. (2007). Does the h-index have predictive power? Proceedings of the National Academy of Sciences of the United States of America, 104(49), 19193–19198.

Hosmer, D.W., & Lemeshow, S. (2000). Applied logistic regression (2. ed.). Chichester, UK: John Wiley & Sons, Inc.

Iglesias, J.E., & Pecharromán, C. (2007a). Comparing h-indices for scientists in different ISI fields. In D. Torres-Salinas & H.F. Moed (Eds.), Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics (Vol. 2, pp. 862–863). Madrid, Spain: Spanish Research Council (CSIC).

Iglesias, J.E., & Pecharromán, C. (2007b). Scaling the h-index for different scientific ISI fields. Scientometrics, 73(3), 303–320.

Imperial, J., & Rodríguez-Navarro, A. (2007). Usefulness of Hirsch's h-index to evaluate scientific research in Spain. Scientometrics, 71(2), 271–282.

Jin, B. (2006). *h*-index: an evaluation indicator proposed by scientist. Science Focus, 1(1), 8–9.

Jin, B. (2007). The AR-index: complementing the h-index. ISSI Newsletter, 3(1), 6.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. Chinese Science Bulletin, 52(6), 855–863.

Kelly, C.D., & Jennions, M.D. (2006). The h index and career assessment by numbers. Trends in Ecology & Evolution, 21(4), 167–170.

Kline, P. (1998). The new psychometrics: Science, psychology, and measurement. London, UK: Routledge.

Kline, R.B. (2004). Beyond significance testing: Reforming data analysis methods in behavioral research. Washington, DC, USA: American Psychological Association.

Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original h-index. ISSI Newsletter, 2(3), 4–6.

Levitt, J.M., & Thelwall, M. (2007). Two new indicators derived from the h-index for comparing citation impact: Hirsch Frequencies and the Normalised Hirsch Index. In D. Torres-Salinas & H.F. Moed (Eds.), Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics (Vol. 2, pp. 876–877). Madrid, Spain: Spanish Research Council (CSIC).

Liu, Y., & Rousseau, R. (2007). Hirsch-type indices and library management: The case of Tongji University Library. In D. Torres-Salinas & H.F. Moed (Eds.), Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics (Vol. 1, pp. 514–522). Madrid, Spain: Spanish Research Council (CSIC).

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

Muthén, L.K., & Muthén, B.O. (1998–2006). Mplus user`s guide (4. ed.). Los Angeles, CA, USA: Muthén & Muthén.

Ravichandra Rao, I.K. (2007). Distributions of Hirsch-index and G-index: An empirical study. In D. Torres-Salinas & H.F. Moed (Eds.), Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics (Vol. 2, pp. 655–658). Madrid, Spain: Spanish Research Council (CSIC).

Rousseau, R. (2006). New developments related to the Hirsch index. Retrieved June 12, 2006, from http://eprints.rclis.org/archive/00006376/

SAS Institute Inc. (2007). SAS 9.1.3. Cary, NC, USA: SAS Institute Inc.

Schreiber, M. (2007). A case study of the Hirsch index for 26 non-prominent physicists. Annalen der Physik, 16(9), 640–652.

Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized Hirsch h-index for disclosing latent facts in citation networks. Scientometrics, 72(2), 253–280.

SPSS for Windows. (2006). Rel. 15.0.1.2006. Chicago, IL, USA: SPSS Inc.

Stevens, J. (1996). Applied multivariate statistics for the social sciences. London, UK: Lawrence Erlbaum Associates.

Timm, N.H. (2002). Applied multivariate analysis. New York, NY, USA: Springer.

Ullman, J.B., & Bentler, P.M. (2004). Structural equation modeling. In M. Hardy & A. Bryman (Eds.), Handbook of data analysis (pp. 431–458). London, UK.

DOI: 10.1002/asi