

BRIEF COMMUNICATION

Do We Need the h Index and Its Variants in Addition to Standard Bibliometric Measures?

Lutz Bornmann and Rüdiger Mutz

Professorship for Social Psychology and Research on Higher Education, ETH Zurich, Switzerland.

E-mail: bornmann@gess.ethz.ch

Hans-Dieter Daniel

Professorship for Social Psychology and Research on Higher Education, ETH, Zurich, Switzerland, and Evaluation Office, University of Zurich, Zurich, Switzerland.

In this study, we investigate whether there is a need for the h index and its variants in addition to standard bibliometric measures (SBMs). Results from our recent study (L. Bornmann, R. Mutz, & H.-D. Daniel, 2008) have indicated that there are two types of indices: One type of indices (e.g., h index) describes the most productive core of a scientist's output and informs about the number of papers in the core. The other type of indices (e.g., a index) depicts the impact of the papers in the core. In evaluative bibliometric studies, the two dimensions quantity and quality of output are usually assessed using the SBMs "number of publications" (for the quantity dimension) and "total citation counts" (for the impact dimension). We additionally included the SBMs into the factor analysis. The results of the newly calculated analysis indicate that there is a high intercorrelation between "number of publications" and the indices that load substantially on the factor Quantity of the Productive Core as well as between "total citation counts" and the indices that load substantially on the factor Impact of the Productive Core. The high-loading indices and SBMs within one performance dimension could be called redundant in empirical application, as high intercorrelations between different indicators are a sign for measuring something similar (or the same). Based on our findings, we propose the use of any pair of indicators (one relating to the number of papers in a researcher's productive core and one relating to the impact of these core papers) as a meaningful approach for comparing scientists.

Introduction

In a recently published study (Bornmann, Mutz, & Daniel, 2008), we examined empirical results on the h index (Hirsch,

2005) and its most important variants (a total of nine different indices). The results of a factor analysis using bibliometric data on recipients of postdoctoral research fellowships from the Boehringer Ingelheim Fonds (BIF; an international foundation for the promotion of basic research in biomedicine located in Heidesheim, Germany) indicated that with the h index and its variants, we are dealing with two types of indices that load on one factor each: One type of indices (e.g., the h index and g index) describes the most productive core of a scientist's output and gives the number of papers in that core (factor: Quantity of the Productive Core). The other type of indices (e.g., the a index and r index) describes the impact of the papers in the core (factor: Impact of the Productive Core). Our conclusion was that the two index types stand for two different performance dimensions of scientists' research output. For evaluation purposes, we proposed the use of any pair of indices (as a meaningful approach for comparing scientists) where the two indices each represent one of the two dimensions.

In evaluative bibliometric studies, the two dimensions quantity and quality of output are usually assessed using the standard bibliometric measures (SBMs) "number of publications" (for the quantity dimension) and "total citation counts" (for the impact dimension) (Hirsch, 2007; van Raan, 2004). Therefore, is there a need for the h index and its variants in addition to these SBMs? To answer this question, we additionally included the SBMs in the factor analysis of our study (Bornmann et al., 2008) and recalculated the analysis.

Methods

Statistical Analysis

Factor analysis is a statistical method "to reduce the dimensionality of the data space in order to discover,

Received June 5, 2008; revised November 20, 2008; accepted November 20, 2008

© 2009 ASIS&T • Published online 2 February 2009 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.21016

visualize, and interpret dependencies among sets of variables” (Timm, 2002, p. 445). Factor analysis provides information on the dimensionality of the structure of the dependencies in a dataset (one-dimensional, two-dimensional, etc.). With regard to the different indices (i.e., the *h* index and its variants) and SBMs (number of publications and total citation counts) for measuring research performance, we examined in this study whether the indices and SBMs are based on two factors (or dimensions). The two-factor solution was the result of our previous study (Bornmann et al., 2008) including the *h* index and its most important variants. In the present study, if both the indices and the SBMs load substantially on two factors, the need for the *h* index and its variants in addition to the SBMs can be questioned. Factor loadings greater than .60 are defined as substantial. According to Velicer and Jackson (1990), only such substantial loadings guarantee high stability of the obtained factor solution regardless of the method of factor analysis (e.g., factor extraction method).

Dataset for the Investigation of the Indices and SBMs

We investigated committee peer review for awarding long-term fellowships to postdoctoral researchers as practiced by the BIF (Bornmann & Daniel, 2005, 2006). According to Fröhlich (2001), managing director of the BIF, applicants that demonstrate excellence in scientific work are selected for the fellowships by the BIF 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 1990 to 1995, with a total of 1,586 papers that they published before 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 Reuters) (citation window: from year of publication to the end of 2001).

Investigated Indices and SBMs

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 (Hirsch, 2005), *g* index (Egghe, 2006), *h*(2) index (Kosmulski, 2006), *a* index (Jin, 2006), *r* index (Jin, Liang, Rousseau, & Egghe, 2007), *ar* index (Jin et al., 2007), and *h_w* index (Egghe & Rousseau, 2008). We also included in our analysis the *m* index, a variant of the *a* index that we propose (Bornmann et al., 2008), and the two SBMs: “number of publications” and “total citation counts.”

Scale Transformation of the Indices and SBMs

As the median of the individual indices and SBMs in part deviates very strongly from the mean and as the Shapiro–Wilks normality test is statistically significant for all of the indices and SBMs, the indices and SBMs 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 or SBMs (*y*, *x*), respectively, is nonlinear and can be described using a power function: $y = f(x) = C * x^{-a}$, where *C* and *a* are constants. Logarithmizing the equation results in a simple linear function: $y = \log_e[f(x)] = \log_e(C) - a * \log_e(x)$. Therefore, all of the index and SBM values of the BIF 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. 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 nonnormality (Curran, West, & Finch, 1996; Muthén & Muthén, 1998–2006).

Results

Using this special variant of ML factor analysis, we tried to find basic dimensions (or factors) that indicate how the indices and SBMs calculated for the BIF applicants cluster (Kline, 1998; Stevens, 1996). In doing so, our set of correlated variables (the indices and SBMs) was transformed into a set of uncorrelated latent variables (dimensions or factors). The factor loading matrix for the factors as well as the indices and SBMs is shown in Table 1. The results of the factor analysis reveal that two factors with eigenvalues larger than 1 explain nearly 95% of the total variance in the covariance matrix of the indices.

The categorization of the indices among the factors by using factor loadings greater than 0.6 (these loadings are in boldface in the table) revealed that the indices can be

TABLE 1. Varimax Rotated Loading Matrix of a Maximum Likelihood Factor Analysis for Nine Log-Transformed Indices and Two Log-Transformed SBMs (*N* = 414).

Index or SBM	Factor 1 (Quantity of the Productive Core)	Factor 2 (Impact of the Productive Core)	Communality
<i>h</i> index	0.93	0.35	0.99
<i>m</i> quotient	0.86	0.29	0.82
<i>g</i> index	0.90	0.31	0.90
<i>h</i> (2) index	0.81	0.51	0.92
<i>a</i> index	0.15	0.99	1.00
<i>m</i> index	0.08	0.98	0.96
<i>r</i> index	0.51	0.86	1.00
<i>ar</i> index	0.50	0.85	0.96
<i>h_w</i> index	0.49	0.87	0.99
No. of publications	0.88	0.10	0.78
Total citation counts	0.51	0.86	0.99
Eigenvalue	4.88	5.47	
Explained variance (%)	44.4	49.7	

Note. Values in boldface denote factor loadings >0.60.

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 [i.e., 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 (i.e., 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 1). These results of the factor analysis are in agreement with the findings of our previous study (Bornmann et al., 2008).

Furthermore, the results in Table 1 show that “number of publications,” with a factor loading of 0.88, loads on Quantity of the Productive Core (Factor 1), and “total citation counts,” with a factor loading of 0.86, loads on Impact of the Productive Core (Factor 2). Both factor loadings are very high, which means that “number of publications” is highly correlated with Factor 1 and that “total citation counts” is highly correlated with Factor 2.

Discussion

Since Hirsch’s first publication of the h index in 2005, a number of h index variants have been proposed (Bornmann & Daniel, 2007, 2009; Bornmann et al., 2008). 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. To identify the factors that mathematically account for the variance in the covariance matrix of the different indices, we calculated an exploratory factor analysis. The results of the analysis indicate that with the h index and its variants, we can assume that there are two types of indices: One type (e.g., h index and g index) describes the most productive core of the output of a scientist and tells us the number of papers in the core. The other type (e.g., a index and r index) depicts the impact of papers in the core.

As in evaluative bibliometric studies, the two dimensions quantity and quality of output are usually assessed using the SBMs “number of publications” (for the quantity dimension) and “total citation counts” (for the impact dimension). In the present study, we tried to assess whether there is a need for the h index and its variants in addition to these SBMs. To do that, we additionally included the SBMs into the factor analysis and recalculated the analysis. The results of the newly calculated analysis indicated that there is a high intercorrelation between “number of publications” and the indices that load substantially on the Quantity of the Productive Core as well as between “total citation counts” and the indices that load substantially on Impact of the Productive Core. The high-loading indices and SBMs within one performance dimension could be called redundant in empirical application, as high intercorrelations between different indicators are a sign for measuring something similar (or the same). Based on our findings, we propose the use of any pair of indicators (one relating to the number of papers in a researcher’s productive

core (e.g., the h index or “number of publications”) and one relating to the impact of the papers in a researcher’s productive core (e.g., the a index or “total citation counts”) as a meaningful approach for comparing scientists.

Our results refer to a dataset from the area of biomedicine (All scientists for whom indices and SBMs were calculated in the present study conduct biomedical research.) For the sake of assessing generalizability, we would very much like to see replication studies conducted using datasets from other scientific fields. If these studies include scientists from several different disciplines, subfield-specific normalized relative indicators in addition to “number of publications” and “total citation counts” (assessed in the present study) should be considered. This seems advisable, given that a recently published study by van Leeuwen (2008) showed that relative indicators correlate significantly lower with the h index than SBMs.

Acknowledgments

We are grateful to three anonymous referees for their helpful comments.

References

- Bornmann, L., & Daniel, H.-D. (2005). Does the h -index for ranking of scientists really work? *Scientometrics*, 65(3), 391–392.
- 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. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9), 1381–1385.
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). 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. *Journal of the American Society for Information Science and Technology*, 59(5), 830–837.
- Bornmann, L. & Daniel, H.-D. (2009). The state of h index research. Is the h index the ideal way to measure research performance. *EMBO Reports*, 10(1), 2–6.
- 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, United Kingdom: Elsevier.
- 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. (2006). An improvement of the h -index: The g -index. *ISSI Newsletter*, 2(1), 8–9.
- Egghe, L., & Rousseau, R. (2008). An h -index weighted by citation impact. *Information Processing & Management*, 44(2), 770–780.
- Fröhlich, H. (2001). It all depends on the individuals. Research promotion—A balanced system of control. *B.I.F. Futura*, 16, 69–77.
- Hirsch, J.E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences, USA*, 102(46), 16569–16572.
- Hirsch, J.E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences, USA*, 104(49), 19193–19198.
- Jin, B. (2006). h -index: An evaluation indicator proposed by scientist. *Science Focus*, 1(1), 8–9.

- 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.
- Kline, P. (1998). *The new psychometrics: Science, psychology, and measurement*. London: Routledge.
- 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.
- Muthén, L.K., & Muthén, B.O. (1998–2006). *Mplus user's guide* (4th ed.). Los Angeles: Muthén & Muthén.
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. London: Erlbaum.
- Timm, N.H. (2002). *Applied multivariate analysis*. New York: Springer.
- van Leeuwen, T. (2008). Testing the validity of the Hirsch-index for research assessment purposes. *Research Evaluation*, 17(2), 157–160.
- van Raan, A.F.J. (2004). Measuring science. *Capita selecta of current main issues*. In H.F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems* (pp. 19–50). Dordrecht, The Netherlands: Kluwer.
- Velicer, W.F., & Jackson, D.N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25(1), 1–28.