RANKING THE SCHOLARLY IMPACT OF NEW JERSEY COLLEGES AND UNIVERSITIES IN THE FIELD OF PSYCHOLOGY

Author:
Michelle T. Guerin

Faculty Sponsor:
John Ruscio,
Department of Psychology

ABSTRACT

With new research published every day, it becomes increasingly difficult to evaluate scholarly impact in a field of study. To address this problem, Hirsch (2005) developed the *h* index to express the impact of a researcher in a way that balances his or her volume and quality of output. Hirsch defined *h* as the largest number such that *h* papers have received at least *h* citations each. This index spurred the development of others that have been applied to measuring the impact of individuals, journals, and research topics. This study uses the *h* index and four of its variants to measure and rank the impact of New Jersey colleges and universities on the discipline of psychology. A principal component analysis showed that the *h* index and the four variants studied here all appear to measure one variable, scholarly impact. They do so in more useful ways than a variety of simpler measures such as the total number of papers published, the total number of citations, or the average number of citations per paper. The rankings produced in this study reveal trends for New Jersey institutions' impact on psychological science, and this project suggests the feasibility and utility of follow-up studies that apply the *h* index and its variants to a broader range of institutions, fields of study, or both.

INTRODUCTION

With new research being published every day, scientists are faced with the challenge of keeping up with the latest findings. A study may prompt a critical review, a follow-up investigation, research seeking to improve on the original study, and so on. In the context of a rapidly expanding body of work appearing in thousands of journals published around the world, how can the influence of individual scholars be evaluated? Because many individuals publish at a prodigious pace, it can be impractical to read all of this work carefully. Moreover, scholarship is increasingly specialized, so even a careful reading may not afford a reliable and valid judgment of the quality of the work. As a result, it is common for a researcher's impact to be examined through the number of peer-reviewed articles that he or she publishes or the number of citations that this work generates. Citation counts and indices based on them are a simple and useful indication of the quantity and quality of scholarship. Citations represent the judgment of relevant experts that work is influential—at least that it is worth citing. Research has shown that citation counts are positively related to peers' ratings of the quality of one's work (Shadish, Tolliver, Gray, & Gupta, 1995).

Hirsch (2005) suggested an index to measure an author's impact using the total number of his or her publications and the citations generated from them. What he named the "*h* index" is defined as the largest number *h* such that *h* papers have received at least *h* citations each. So, for example, given the rank-ordered citations counts 10, 8, 7, 5, 3, 2, 1, 0, 0, 0 for a set of 10 articles, the value of *h* for this set would be 4. The fourth article was cited at least four times, and the fifth article was cited fewer than five times. This is depicted in Figure 1 (top graph), which shows that *h* is the length of the largest square that fits within an array of citations by articles.

The *h* index revolutionized the study of scholarly impact because of its conceptual simplicity and its ability to communicate, in a single number, an indication of the volume and quality of a researcher's output. Hirsch (2007), in a follow-up study of the validity of the *h* index, found that researchers' future

achievement was better predicted by the *h* index than by the number of papers published (*N*), the total number of citations received (*C*), or by the mean number of citations per paper (*M*). The *h* index has also been found to be robust to missing articles and papers (Rousseau, 2007) because they typically receive few citations and would not increase *h*.

This index was originally introduced in the physical sciences and soon spread to social sciences. It has been used successfully to rank individuals (Hirsch, 2005), journals (Braun, Glanzel, & Schubert, 2006), and research topics (Banks, 2006). Many variations of the *h* index theme have been suggested as ways to improve its ability to capture both the quantity and quality of scholarly work. Among these are *g*, tapered *h* (h_t), and a size-adjusted *h* (h_m).

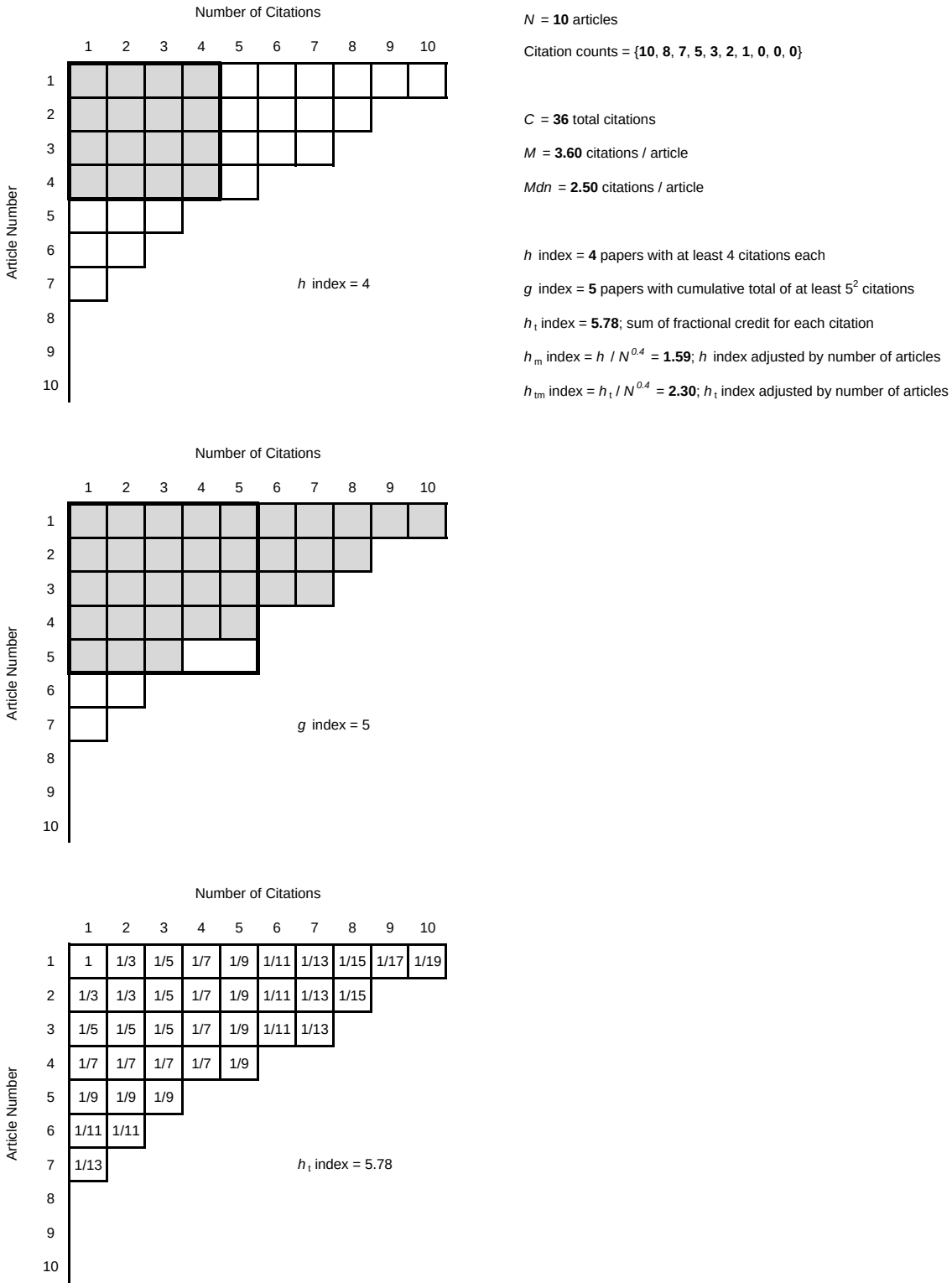

Figure 1. Illustrating the citation-based indices of scholarly impact.

The *g* index was developed by Egghe (2005) to award more credit than the *h* index for highly cited papers. The *g* index is defined as the "largest number such that the top *g* articles received (together) at least g^2 citations" (Egghe, 2005, p. 131). Graphically, *g* is the length of the largest square such that the

first g articles can fill the square. The simplest way to illustrate the g index is by using the same data and keeping a cumulative total, as shown in Figure 1 (middle graph). Citations that lie outside the square on the right are allowed to fill in missing spaces within the square. In other words, the g index allows a large number of citations for one or more papers to offset a lower number of citations for other papers. If the pattern of citations across articles formed a perfect square, g would equal h. If some of the citations are located to the right of the h square, g can award additional credit. Thus, for all possible sets of citation counts, $g \geq h$. For the illustrative data set introduced earlier, $g = 5$ (recall that $h = 4$ for these data).

The next variation of the h index is the tapered h (h_t), which assigns a fractional value to each citation received (Anderson, Hankin, & Killworth, 2008). Thus, unlike h and g, h_t need not be a whole number. Figure 1 (bottom graph) illustrates the fractional credit assigned to each citation; the pattern extends as far in each dimension—citations and articles—as required to encompass all citations. If the pattern of citations across articles formed a perfect square, h_t would equal h. If a single article is cited once ($h = 1$), the lone citation receives one point ($h_t = 1$). If two articles are each cited twice ($h = 2$), the first citation receives one point and the next three citations received one-third point each, for a total of $h_t = 2$. If three articles are each cited three times ($h = 3$), the first citation receives one point, the next three receive one-third point each, and the remaining five receive one-fifth point each, so $h_t = 1 + 3(1/3) + 5(1/5) = 3$. For any citations lying outside of the h square, h_t awards additional credit. Thus, for any possible set of citation counts, $h_t \geq h$. For the illustrative data set introduced earlier, $h_t = 5.78$.

To compare institutions rather than individuals, it is useful to control for the size of the institution. Otherwise, indices such as h, g, or h_t tend to confound the size of the institution with the impact of its scholarship because, all else being equal, simply publishing more articles increases scores on these indices. Molinari and Molinari (2008) proposed the h_m index to compare the impact of institutions of disparate sizes or across disciplines that have a significant difference in volume of publications. The h_m index is a transformation of h using the following formula:

$h_m = h / N^{0.4}$, where N = number of publications.

The value of h_m for the running example is $4 / 10^{0.4} = 1.59$.

A final variant of the h index is simply a combination of h_t, which awards fractional credit for every citation, and h_m, which adjusts for the number of articles published. This new index is labeled h_{tm} and it is calculated by substituting h_t in the place of h in the h_m formula, as follows:

$h_{tm} = h_t / N^{0.4}$

The value of h_{tm} for the running example is $5.78 / 10^{0.4} = 2.30$.

The present study seeks to rank colleges and universities in New Jersey according to their scholarly impact in the field of psychology. After calculating traditional measures including the total number of articles, total number of citations, and mean number of citations per article, the h index and its four variants will also be calculated. The goals of this investigation are twofold. First, data will be collected to provide insight into the influence that each of these institutions has in the realm of psychological science. How do these schools compare to one another? Second, analyses will be performed to examine the performance and utility of the many available measures of scholarly impact. Does it make much difference which index is used to rank schools, and if so how can someone select an appropriate index?

METHOD

A list of 24 New Jersey colleges and universities with psychology departments was obtained from the following web site: http://www.uscollegesearch.org/new-jersey-psychology-colleges.html. The three Rutgers University campuses (Camden, New Brunswik/Piscataway, and Newark) were combined into a single institution. Because Centenary College did not have any publications listed in the database and time frame used for this study (see below), it did not receive any scores and was dropped. This brought the final sample to 21 institutions.

Citation data were obtained from the PsycINFO database. Searches were performed using each institution's name in an author affiliation query. For each article that was retrieved, "Times Cited in Database" was recorded. Data were restricted to peer-reviewed journal articles (which excludes dissertations) published between January 2003 and December 2008. This time frame was chosen for two reasons. First, the aim of this study was for a citation index to reflect the recent impact that departments have had, and not their historic impact. Second, previous studies suggest that a five-year time frame is neither too short nor too long for measuring articles' effect (Adair & Vohra, 2003; Moed, van Leeuwen, & Reedijk, 1999). The six years of the present study mean that articles published as far back as 2003 have had at least a full five years to be cited. True, more recently published articles have not had as much time for citation, but this is an unavoidable cost of evaluating contemporary (rather than historical) scholarly impact.

The PsycINFO database is updated frequently, so to control for the effect of these updates on institutions' scores, all searches were conducted within the three-week period from March 16, 2009 to April 5, 2009. Institutions were sorted randomly prior to our performing searches for one institution at a time. In addition to recording the total number of citations for each article, it was noted whether or not the author affiliated with a target institution was the first author. This was done to consider whether or not a restriction to first-authored publications would significantly affect the ranking of institutions.

RESULTS

Before ranking institutions by scholarly impact, we calculated Spearman's rank-order (r_S) correlations to compare the results using data for first-authored articles only to the results using data for all articles. These correlations ranged from $r_S = .615$ to .989 (see Table 1), and all were statistically significant ($p < .001$) using a two-tailed test. The value of $r_S = .615$, for the *Mdn*, was an outlier on the low end because most institutions were tied at *Mdn* = 0 citations per article (meaning that most articles were not cited even once). Setting aside this value, the next-lowest correlation was $r_S = .828$, which suggests very little difference in rank orders. The fact that both versions of each index were so highly correlated indicates that authorship position did not exert a strong influence on the rank-ordering of institutions. All subsequent analyses were based on data for all articles to present a more holistic picture of the institutions.

Using all articles, the indices of scholarly impact were calculated for each institution (see Table 2). The number of articles ranged from $N = 2$ to 1,436 in a highly skewed distribution. After Rutgers University ($N = 1,436$), Princeton University had the second largest number of articles ($N = 714$) and all other institutions had far fewer ($N \leq 258$). Across all 21 institutions, data were recorded for a total of 3,404 articles. The number of citations received by each institution ranged from $C = 0$ to 5,658. By this measure, Rutgers University ($C = 5,036$) and Princeton University ($C = 5,658$) surpassed all others ($C \leq 436$) by a substantial margin.

Measure	Correlation
N	0.996
C	0.989
M	0.893
Mdn	0.615
h	0.968
g	0.969
h_t	0.983
h_m	0.828
h_{tm}	0.918

Table 1. Rank-order correlations between index values calculated for first-authored articles and all articles.

The mean number of citations per article ranged from *M* = .00 to 7.91. Though the ranking of institutions by *M* was similar to the rankings by *N* or *C*, there were notable exceptions. For example, the College of Saint Elizabeth had fewer articles than any other institution (*N* = 2), yet its *M* of 1.50 citations per article was ranked 11th out of 21. Clearly, *M* can yield quirky results when the number of articles ranges so low.

Institution	*N*	*C*	*M*	*Mdn*	*h*	*G*	h_t	h_m	h_{tm}
Princeton University	715	5,658	7.91	2	35	58	61.96	2.52	4.47
Rutgers University	1,436	5,036	3.51	1	28	39	52.00	1.53	2.84
Drew University	21	88	4.19	0	5	9	7.44	1.48	2.20
The College of New Jersey	144	303	2.10	0	10	13	15.58	1.37	2.13
Rider University	76	189	2.49	1	6	11	11.80	1.06	2.09
Montclair State University	219	436	1.99	0	10	13	16.89	1.16	1.96
Rowan University	72	136	1.89	0	7	9	10.63	1.26	1.92
Fairleigh Dickinson University	98	157	1.60	0	7	9	11.22	1.12	1.79
Seton Hall University	117	182	1.56	0	6	10	11.14	.89	1.66
Kean University	45	67	1.49	0	4	6	7.34	.87	1.60
Richard Stockton College	258	300	1.16	0	9	11	14.37	.98	1.56
Monmouth University	69	81	1.17	0	4	8	7.88	.74	1.45
William Paterson University	60	70	1.17	0	4	6	7.38	.78	1.44
Ramapo College of New Jersey	31	51	1.64	0	3	6	5.13	.76	1.30
Felician College	5	6	1.20	0	2	2	2.40	1.05	1.26
Saint Peters College	9	7	.78	0	2	2	2.54	.83	1.06
College of Saint Elizabeth	2	3	1.50	1.50	1	1	1.53	.76	1.16
Caldwell College	7	5	.71	0	2	2	2.20	.92	1.01
Georgian Court University	8	3	.38	0	1	1	1.67	.44	.72
N.J. City University	9	3	.33	0	1	1	1.53	.42	.64
Bloomfield College	3	0	.00	0	0	0	.00	.00	.00

Table 2. Indices of scholarly productivity for each institution. Entries are ranked by scores on the size-adjusted tapered *h* index (h_{tm}).

The median number of citations per article was *Mdn* = 0 for most departments; the exceptions were two institutions with *Mdn* = 1, one with *Mdn* = 1.50, and one with *Mdn* = 2. Though some have argued that *M* is influenced too much by outliers and skew and suggested that *Mdn* is a more appropriate measure, most articles are not cited much if at all. This results in a tendency for *Mdn* to be very low and yields many tied scores of 0, supporting the use of *M* rather than *Mdn* if one wants to rank institutions.

Based on the conventional measures reviewed so far, based on the total number of publications and citations, the top five institutions are Princeton University, Rutgers University, Montclair University, The College of New Jersey, and The Richard Stockton College of New Jersey.

Next, a matrix of rank-order correlations (r_S) was created to examine relationships between the citation-based indices of scholarly impact. Table 3 shows that all indices were highly correlated, and statistically significantly so ($p < .001$) using a two-tailed test. Correlation coefficients ranged from r_S = .802 to .989, suggesting that the rank-ordering of most institutions would remain unchanged across indices. To evaluate which of these indices provides a better measure of the core construct of scholarly impact, a principal-components analysis was conducted, with the solution constrained to a single component representing scholarly impact. This single component captured 93.2% of the variance on these variables. The results of this analysis showed the following loadings on the component: *h* = .960, *g* = .967, h_t = .947, h_m = .852 and h_{tm} = .933. This analysis shows that the three indices that do not adjust for the number of articles published—*h*, *g*, and h_t—had higher loadings than the two indices that involve

such an adjustment—h_m and h_{tm}. It is noteworthy that the h_{tm} index exhibited a loading nearly as strong as the former three indices even though it adjusted for the number of articles, suggesting that it may be especially useful for comparing institutions that vary in the amount of research produced.

	h	g	h_t	h_m
g	0.981			
h_t	0.985	0.989		
h_m	0.851	0.831	0.802	
h_{tm}	0.917	0.929	0.903	0.888

Table 3. Rank-order correlations between citation-based indices of scholarly impact.

Institutions' scores on these indices were presented in Table 2, in descending order by the h_{tm} index. Using any of these five indices, Princeton University is ranked first and Rutgers University, second. Beyond these two large, research-intensive institutions, the rankings depend to some extent on the index selected. Using the h, g, or h_t indices places Montclair State University and The College of New Jersey in the third and fourth positions. However, using the indices that adjust for the number of articles (h_m or h_{tm}), Drew University surpasses both of these institutions and Rider University edges out Montclair State but not The College of New Jersey. As the correlations in Table 3 demonstrate, most of the rankings would change relatively little across indices. Because these were rank-order correlations, the fact that Rutgers and Princeton were outliers in skewed distributions does not exert undue influence on the correlation coefficients.

Given the uniformly high correlations, any of these indices should be useful. However, one should bear in mind that they address somewhat different questions. For example, if one wanted to know which institutions' research exerts the greatest total impact on psychology, then either the h, g, or h_t index would be most appropriate. On the other hand, if one wanted to know which institutions' faculty exert the greatest impact on psychology *given* the total quantity of papers that are published in peer-reviewed journals, then either the h_m or h_{tm} index would be most appropriate. In other words, a meaningful distinction can be made between total scholarly impact and scholarly impact given the amount of research that is published. One might compare research-focused institutions (e.g., Rutgers University) using the former type of metric, and institutions that expect faculty to strike a balance between teaching and scholarship (e.g., The College of New Jersey) using the latter type.

DISCUSSION

According to their very high loadings on a single component, the indices that have been developed over the past four years seem to measure the core construct of scholarly impact. Although the g index had a slightly higher loading in the principal component analysis, it is not necessarily the best measure for ranking departments. Each measure has its strengths and weaknesses.

The h index provides meaningful and accurate results, and at the same time is easy to calculate and understand (Saad, 2006). This is the only index that is readily available on the Thompson ISI ("Web of Science") database, making it especially convenient for evaluating the scholarly output of scientists. The h index is better at predicting future achievement than the total number of papers published, total number of citations, and mean number of citations per paper (Hirsch, 2007). However h is also limited since it does not account for the citations that lie outside of the defining square and so a few highly cited articles and several less-cited ones end up being ignored in this calculation. Moreover, h and g yield whole numbers only, which can lead to many tied scores. By contrast, the other indices provide fractional scores so there are more distinct values that can be attained and less likelihood of ties.

The *g* index received the highest loading on the analysis, suggesting that it improves on *h* by better awarding credit to the most highly cited articles. However, it still does not award credit for all the less cited articles that *h* also ignores and, as noted above, it yields many tied scores.

The h_t index not only accounts for the highly cited articles, but also awards credit to every single publication and citation, including the very seldom cited publications that both *h* and *g* ignore. The limitations with h_t are that it is not as simple and straightforward to understand and that it is more complex to calculate. For example, *h* can be accessed through a database or easily obtained by proceeding through a list of rank-ordered citation counts until a value falls below its position on the list. Calculating h_t, on the other hand, requires awarding credit using a matrix of fractional values that would be very cumbersome to generate and total without a computer. Calculating *g* is of intermediate complexity, requiring the summation of citation counts and comparison to the cumulative number of articles squared. Despite its complexity, there are three major strengths to h_t that suggest its utility in research and practice. First, unlike *h* or *g*, it awards some credit for every citation while still balancing the quantity and quality of output. Second, it assesses the core construct of scholarly impact with excellent validity, as evinced by its strong loading in the principal components analysis. Third, it is very unlikely to yield tied scores.

The h_m index shares the same limitations with *h* by ignoring citations "out of the box" but it generates more distinct scores (decimal values) and adjusts for differences in scholarly output in a way that can facilitate comparisons at aggregate levels of analysis (e.g., when the unit is not an individual scholar but an institution or journal). This is also a fairly straightforward index to calculate, and therefore remains simple to understand and explain. It performed very well in the principal components analysis.

The h_{tm} index, which was created for this study, combines the strengths of the h_t index—awarding credit for all citations—and the h_m index—taking into account the number of articles. This, too, performed very well in the principal-components analysis. In light of the strengths and weaknesses of each index, it seems that h_{tm} may be the measure of choice when the unit of analysis is broader than the individual scholar.

Whereas the *h* index is provided by one major database of published research, the other indices must be calculated using retrieved citation counts. That means not only that more effort goes into calculating these indices, but there is more room for human error. In addition, every index is limited by the database used. Journal coverage varies across databases, and it might be particularly challenging to retrieve all citations for interdisciplinary work that crosses the boundaries of databases covering traditional domains. Rousseau (2007) addresses concerns about missing articles and citations by arguing that regardless of the limitations of the database used, *h* is robust to missing entries. The assumption underlying this argument is that missing entries correspond to articles with relatively low citation counts, entries that would not affect *h*. Presumably, each of the other indices would be less robust to missing entries, since they are or can be influenced by articles receiving few citations.

Moreover, the databases used in any particular study, such as PsycINFO for the present one, are constantly updated. Thus, the number of articles listed and the citation counts for each article constantly change. In this study, data retrieval was performed within a three-week time frame and done by randomly sorting institutions so these updates had no systematic impact on any scores. Nonetheless, the scores presented in Table 2 represent a temporal snapshot. While it seems exceptionally unlikely that Princeton University and Rutgers University would fail to occupy the first and second positions for scholarly impact in psychology among New Jersey colleges and universities, the rankings for many of the other institutions could change were the time frame extended forward or backward.

In light of the attention given to citation counts in evaluating scholarly impact, researchers have begun to scrutinize the information value of the citations. What do those citations actually represent? Does authorship position matter, and how many of the citations represent self-citations? The present study suggests that restricting analysis to first-authored publications, rather than including all publications, would not change the results much. After three studies on why certain articles are cited, Shadish, Tolliver, Gray, and Gupta (1995) found that citations were never made for personal reasons, such as citing oneself to boost citation count or doing something similar for a friend or colleague. They

also found that citation counts are positively related to peer ratings of research quality. Moreover, self-citations are not always invalid. Often it is appropriate, even necessary, for authors to refer to their own original work, particularly when one engages in a successful program of research. Regardless of why the self-citations occur, some researchers argue that because most authors do cite themselves, it is not necessary to control for self-citations (Cronin & Meho, 2006). Other researchers indicate that by sampling from a broad enough period (five years was suggested), any adverse effect of including self-citations is mitigated (Moed, van Leeuwen, & Reedjik, 1999).

In addition to evaluating the merits of a variety of measures of scholarly impact, including the new h_{tm} index that appears to combine the advantages of the h_t and h_m indices effectively, the second major goal of this study was to rank the impact of New Jersey colleges and universities on the discipline of psychology. The citation-based indices afford quantitative comparisons between institutions, each summarizing in a single number what impact that school has on the field of psychology as reflected by the last six years of peer-reviewed publications. One should note that, because the searches were performed by institution and not by department, the articles and citations that were retrieved do not indicate the impact of faculty in psychology departments. Articles may have been published by members of any department in the schools and still had an impact on psychology. Fields whose scholarship overlaps with that of psychology include biology, economics, sociology, philosophy, and more. Whatever articles that were retrieved in the PsycINFO searches for each institution were counted towards its scoring regardless of the departmental affiliations of the authors.

Additional research could be done to look specifically at what members of the psychology departments themselves are generating and to examine the differences between the scores in the present study and the scores when the impact of psychology departments are assessed more specifically. This would be much more laborious, as it adds an additional step to the data collection that requires close inspection of the affiliation of each author, and this information is not always apparent from the entries in databases such as PsycINFO.

Having demonstrated that it is feasible to assess scholarly impact using citation-based indices that can be calculated from data retrieved on a very large scale—the present study included all articles published in the past six years, for a total of 3,404 papers—the logical next step would be to expand this type of research. The indices studied here could be calculated to rank the impact on psychology of all colleges and universities in the United States, or internationally. Another way to expand this type of research would be to access databases that catalogue scholarship in other disciplines and provide citation counts. This line of research could prove useful for students trying to decide which schools to apply to for undergraduate or, perhaps even more important, graduate study. Institutional rankings of research productivity could also be used to evaluate the facilities and infrastructure available in support of grant applications or for any other purpose for which academic recognition or visibility is pertinent. Institutions or programs could evaluate themselves relative to those who score better, worse, or at a comparable level to provide quantifiable goals for institutional or disciplinary progress. Indeed, the availability of citation counts for scholarly output and the development of indices based on this raw data open many interesting and important avenues of research. The present study represents a test case that demonstrates the feasibility of such an approach and suggests its value when performed on even larger scales.

REFERENCES

Adair, J., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist*, *58*, 15-23.

Anderson, T. R., Hankin, R. K., & Killworth, P. D. (2008). Beyond the Durfee square: Enhancing the *h*-index to score total publication output. *Scientometrics, 76,* 577-588.

Banks, M. G. (2006). An extension of the Hirsch index: Indexing scientific topics and compounds. arXiv:physics/0604216v2 [physics.soc-ph]. 14 Sep 2006.

Braun, T., Glanzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, *69,* 169–173.

Cronin, B., & Meho, L. (2006). Using the *h*-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology, 57,* 1275-1278.

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

Hirsch, J. (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*, 16569-16572.

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

Moed, H. F., van Leeuwen, T. N., & Reedijk, J. (1999). Towards appropriate indicators of journal impact. *Scientometrics, 46*, 575-589.

Molinari, J. F., & Molinari, A. (2008). A new methodology for ranking scientific institutions. *Scientometrics, 75*, 163-174.

Rousseau, R. (2007). The influence of missing publications on the Hirsch index. *Journal of Informetrics*, 1, *2-7.*

Saad, G. (2006). Exploring the *h*-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics, 69*, 117-120.

Shadish, W. R., Tolliver, D., Gray, M., & Gupta, S. K. (1995). Author judgments about works they cite: Three studies from psychology journals. *Social Studies of Science, 25,* 477-498.