Journal of the American Society for Information Science and Technology

QUANTIFYING SCHOLARLY IMPACT: IQp VERSUS THE HIRSCH h

Journal:	*Journal of the American Society for Information Science and Technology*
Manuscript ID:	JASIST-2007-05-0157.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Antonakis, John; University of Lausanne, Faculty of Business and Economics Lalive, Rafael; University of Lausanne, Faculty of Business and Economics
Keywords:	bibliometrics < informetrics < measurement < quantitative research < (research and analytic methods)

Abstract

Hirsch's (2005) *h* index of scholarly output has generated substantial interest and wide acceptance because of its apparent ability to quantify scholarly impact simply and accurately. We show that the excitement surrounding *h* is premature for three reasons: *h* stagnates with increasing scientific age; it is highly dependent on publication quantity as well as field-specific citation rates. Thus, it is not useful for comparing scholars across disciplines. We propose the scholarly Index of Quality and Productivity (*IQp*) as an alternative to *h*. The new index takes into account a scholar's total impact and also corrects for field-specific citation rates, scholarly productivity, and scientific age. The *IQp* accurately predicts group membership on a common metric, as tested on a sample of 80 scholars from three populations: (a) Nobel winners in Physics (n=10), Chemistry (n=10), Medicine (n=10), and Economics (n=10), and towering Psychologists (n=10), and scholars who have made more modest contributions to science including randomly selected (b) fellows (n=15) and (c) members (n=15) of the Society of Industrial and Organizational Psychology. The *IQp* also correlates better with expert ratings of greatness than does the *h* index.

Introduction

Hirsch's (2005) *h* index, a measure of a researcher's output, has created quite a stir in the scientific world (see Ball, 2005) and is seen as a useful and simple measure of output and quality (Bornmann & Daniel, 2005; Cronin & Meho, 2006; Oppenheim, 2007). Hirsch (p. 1) defined *h* as "the number of papers with citation number higher or equal to *h*." Thus, a researcher who has an *h* of 13 has 13 papers with at least 13 citations each. Hirsch suggests that the higher the *h* the more of a broad and consistent impact a scholar has had on the field. Such has been the interest in *h* that Thomson's ISI Web of Knowledge (and in particular the Web of Science WoS) now reports this index in its "Citation Report" function. Is this wide acceptance of *h* substantiated?

Limitations of *h*

The *h* index is potentially useful for quantifying the impact of scholarship, and hence its quality. This index could be used as a probabilistic cue in the context of hiring or promotion decisions, grant allocations and the like--particularly because it is supposed to be unbiased. It discounts unique, highly cited papers (i.e., "one-hit wonders", Cronin & Meho, 2006) and rewards a wide and consistent body of work (Hirsch, 2005). The *h* index is attractive because of its simplicity; because of this simplicity, however, it is inherently limited (for detailed critiques refer to Glänzel, 2006a and van Raan, 2006). As we indicate below, *h* does not provide a good indication of a scholar's impact on the field because it is strongly dependent on output and academic discipline. It is thus not useful as a gauge of quality or for cross-disciplinary comparisons. Although the latter limitation is acknowledged by Ball (2005) who called for further research in terms of how to measure impact in a comparable manner across disciplines, the *h* index per se, and as reported by WoS, is limited because it fails to control for the following:

1. *The total impact of an author* (the WoS *h* fails to consider total cites from all sources). That is, scholars receive citations for books, book chapters, working papers, or journals that might not be indexed in WoS (i.e., in the General Search Function), and will thus not be reflected in WoS's calculation of *h.* Writing in what we will term "peripheral" outlets is widely practiced in the social sciences but less so in the exact sciences. Citations from peripheral publications will actually appear in the "Cited Reference Search" of WoS if they have been cited by journal articles indexed in WoS. A scholar's impact on the field is obviously related to the quality of their work and whether it has been cited by others (and citations are generally accepted to be the best way to quantify impact, see Bergh, Perry, & Hanke, 2006). Thus, ignoring peripheral citations underestimates a scholar's true worth. These citations are not captured by the WoS *h* (though a manually-generated *h* would capture these citations). However, these citations would not be ignored by peers who might evaluate the impact of a scholar on a field.

Furthermore, the *h* index severely underestimates total impact of an author, particularly if the author has had one or several very highly cited pieces (the latter is often the case of highly influential authors). Although *h* was designed, in part, to discount authors who might have "struck it lucky" once or twice--by definition unlikely events--it would be very unusual for an author to have only one or a couple of lucky strikes and still be in the academic game. It is easy to imagine the case of doctoral students who were included as coauthors on a highly cited paper coauthored with their doctoral supervisor(s) and then never went on to become career academicians--because such cases are irrelevant, this apparent strength in *h* is not utile.

In the unlikely event that an author hit the jackpot only once or twice and continued to produce output--which will be the case if they remain in the academic field--then a correction must be made for this extra output, which will not be as well cited as the lucky strikes, as well as for each additional year this individual is in the publishing game (as discussed below).

2. *The academic age of the author*, that is, for how many years the author has been active in the publishing game (*h* ignores cites received per year). A bibliometric index should appropriately account for academic age. Given how *h* has been constructed, we would expect it to increase with age (as the author publishes and receives cites). However, after some time in the publication game, it will stagnate and it will become increasingly difficult to increase *h* because the next paper that the author publishes must obtain h_{old} + 1 citations to increase *h*. For example, assume that an author has an *h* of 30. For the next paper to increase *h* to 31, that *particular* paper must receive 31 citations. Imagine that during a certain period of time the author receives 500 citations dispersed across all her papers but does not receive 31 citations on the *particular* paper. In this case, *h* will not budge. The author though, has continued to have an impact on the field and this impact will result in a zero increase in *h.* As the author gets older and generates further output, it will thus become increasingly more difficult to increase *h*, until *h* stagnates.

H1: the relation between h and age is nonlinear such that it increases rapidly at low levels of age and then flattens out at high levels of age (i.e., the quadratic effect is negative).

Also, *h* can only increase or stay constant; however, it can never decrease. As such it does not measure current performance. If authors have been inactive in the field for a long period of time and their work fails to be cited, this information must be included in an index. Thus, authors who receive additional cites per year should be rewarded whereas authors who do not receive additional citations per year should pay a penalty. If the author's impact on the field atrophies then the bibliometric index should reflect this decrease of influence. It is evident that *h* does not have these desirable properties.

3. *The number of papers the author has written* (and thus *h* ignores cites received per paper published). The *h* index does not consider the number of papers a scholar has published in its estimation of a scholar's impact and seems to reward quantity more than quality of output (although it does still acknowledge quality to a certain degree). Thus, a particular scholar, who has published many papers that have received a fair amount of citations (e.g., an *h* of 15 with a total of 1,000 citations, will appear to have (a) the same impact on the field as someone who published the same amount of papers but who has several highly cited papers (e.g., an *h* of 15 papers and 2,000 citations) and (b) a bigger impact on the field as compared to someone who published fewer papers that have each been highly cited (e.g., the case of an author with an *h* of 12, who has 3,000 citations. To accurately judge the quality of research output, scholars should be rewarded for additional citations received per paper published (and pay a penalty for not having received citations for additional papers published).

We are therefore suggesting that *h* depends on quantity of output as its distal predictor--in fact, as suggested by Egghe and Rousseau (2006) and demonstrated by van Raan, 2006, h depends on $\sqrt{number\ of\ \ papers}$. Of course, temporally speaking, output will first predict citations, which will in turn predict *h.* Thus, controlling for all unobserved academic discipline effects (due to field-specific differences in citation raters, see below), $\sqrt{number\ of\ \ papers}$ should be a significant predictor of *h*, as depicted in the following causal model: $\sqrt{number\ of\ \ papers} \rightarrow$ number of cites $\rightarrow$ *h*.

H2: The effect of $\sqrt{number\ of\ \ papers}$ *on h is mediated by number of cites*

4. *Citation rates across fields*. Given its discipline specificity, the *h* index is not interpretable across fields and will thus not be able to predict group membership of individuals who are similar, in terms of their publishing abilities, but who are from different fields. In some fields, average citation rates are very high, as evidenced by higher average journal impact factors because of a large amount of multiple authors (Batista, Campiteli, Kinouchi, & Martinez, 2006), substantially smaller article length (because the study might not require much theory development, among other reasons), and so forth. Thus, *h* will be highly accentuated in some disciplines and highly attenuated in others, making cross-disciplinary comparisons impossible.

H3: h cannot predict group membership of individuals known to be drawn from a similar population of individuals (e.g., Nobel winners) but who are from different disciplines.

Below we present a new index, the *IQp* (Index of Quality and Productivity--a professorial IQ of sorts!) and the rationale for its construction. We describe how the index is tested and calibrated with a sample of scholars and demonstrate how it performs in comparison to the *h* index (using data from ISI's Web of Knowledge). Finally, we discuss the utility of *IQp* in cross-disciplinary comparisons and provide indications of what might constitutes a "good" *IQp* for judging scholarly output and quality, be it for promotion decisions or for winning Nobel prizes.

The *IQp*

As indicated in the above discussion, there is a need to control for the weaknesses of *h* into an intuitive index that is simple to calculate and better captures the true caliber of a scholar. We propose an alternative to the *h* index. As with the intended purpose of *h* (Glänzel, 2006b), the *IQp* measures a scholar's impact along two dimensions: quality (i.e., citations) and productivity (i.e., output). Both factors in synergy (i.e., the interaction of the two components) is what will predict scholarly achievement.

H4: Scholarly achievement depends on the multiplicative association of quality and productivity.

The first component of the *IQp* refers to total number of citations. Citations have to be adjusted, however, for differences in scientific age, number of papers, and field specific citation rates. An intuitively appealing way to combine these three factors is to calculate the expected total number of citations the scholar's work would receive if it were of average quality in the field (cf. van Raan, 2006). We label this component estimated citations (est.citations). The ratio of actual citations to est.citations is a measure of the quality of a scholar's work, expressing how much more often a scholar is actually cited compared to how often she or he would be cited on average.

The second component of *IQp* is a measure of her or his productivity[i]. This component should also be part of an index of scholarly impact because scholars who have been producing more have more impact (all else equal, see below for a definition of productivity). We therefore define *IQp* = Quality * Productivity

$$= \underbrace{\frac{citations}{est.citations}}_{Quality} * \underbrace{papers * \frac{est.citations / paper}{papers + \frac{est.citations}{papers}}}_{Productivity}$$

$$= \frac{citations}{papers + \frac{est.citations}{papers}}$$

Note, estimated citations (est. citations) are the estimated citations the work of the scholar would have received if it were of average quality in her or his field. *IQp* therefore reflects both the quality of an author's work (measured by additional impact of the author's work compared to the average impact in his or her top three fields (i.e., subject areas) and his or her total productivity.

Total productivity is measured in "adjusted papers" reflecting the number of papers published and the impact of the average paper in the author's top three subject areas (i.e., fields, refer to Point 1 below regarding the correction factor). The idea is to bring papers that are published in fields with low impact per paper on the same scale as papers in fields with high impact per paper. We measured impact per paper by relating estimated citations per paper to papers + estimated citations per paper. This ratio is strictly between 0 and 1 and monotonically increasing in the estimated citations per paper. This implies that in comparing two authors that have published the same number of papers, the author contributing to a field with higher impact per paper will end up with a higher number of “adjusted papers”. Moreover, authors publishing few papers contributing to a high impact field can end up having a higher number of “adjusted papers” than do authors with many papers who contribute to a low impact field. Thus, the productivity component of *IQp* measures the number of papers published adjusted for impact per paper in the author’s field.

Estimated citations are:

$$est.citations = \frac{c * age * (papers + 1)}{2}$$

Note, *c* is a correction factor (discussed below) reflecting the citations an average article receives in a particular subject area. This above expression can be obtained by assuming that the author publishes the actual number of papers in equally spaced intervals over her or his scientific age receiving the average citations per paper and year in her or his field[ii]. The estimated average number of citations per paper is therefore

$$\frac{est.citations}{paper} = \frac{c * age * (papers + 1)}{2 * papers}$$

Thus, a young scholar who has published one paper one year after getting his or her Ph.D. in a field where the average article is cited once per year receives 1 est. citations per paper.

Thus, the simplified formula for calculating *IQp* is:

$$IQp = \frac{citations}{papers + \frac{age * c * (papers + 1)}{2 * papers}}$$

This ratio relates the actual output of a scholar to the total inputs needed to produce the standard number of citations in his or her field. We now explain each component of the index, and how to obtain it, in more detail:

1. *c* = Correction factor based on the aggregate journal impact factor (weighted) of the top three subject categories in which the person has been cited[iii]. The journal impact factor reflects the average number of times all articles that have been published in the previous two years have been cited in the current year. This is the correction factor that we introduced to put individuals from different disciplines--where citation rates and research output vary widely--on an equal scale (cf. van Raan, 2006). Recall, the reason why citation rates vary in different disciplines is because of cultural norms and scientific practices relative to number of authors on articles, length of articles, and as well as other reasons. Information to construct *c* is easily obtained from the Cited Research Search function in WoS (after selecting all papers that cite the author and clicking on the “finish search” function, the “Analyze Results” function reports various indicators, including the “Subject Areas” in which the author has been cited)[iv].

The first three subject areas in which the author is cited will provide a good approximation of the major fields in which the author is actively receiving most of their citations, and thus simplify the information need to calculate a scholar’s *IQp*[v]. To use an example, Philip M. Podsakoff--a prominent industrial psychologist--received most of his citations in the following areas (in order of magnitude): Management, Applied Psychology, and Business. To obtain the average citation rates for journals in particular subject categories, one has to use Thomson’s Journal Citations Reports and request the “Journal Summary List” information. This function indicates that the aggregate impact factor for all journals in the three respective fields is 1.042, 1.109, and .932. Weighting these impact factors by order of importance gives Podsakoff a *c* of (1.042*3+ 1.109*2 + .932*1)/6 = 1.046.

2. Citations = Total cites from all publications. This is easily tallied by using the Cited Reference Search of WoS. Note that the total cites is sum of the “Times Cited” column, that must

be counted manually (or simply copied in Excel and counted automatically), otherwise citations are undercounted by WoS (i.e., WoS counts how many times a paper cited an author irrespective of whether the paper has cited more than one of an author's papers). To use an example Podsakoff has 3650 cites when counted manually (but 2505 based on the Cited Research Index search).

Note that Kacmar and Whitfield (2000) have suggested that simply counting citations does not correctly indicate the impact of an author's work on others because counting citations does "not tell us whether the cited work is simply mentioned as a reference or if it is more central to the core of the citing study" (p. 392). They content-analyzed all articles citing a sample of articles in two management journals (Academy of Management Journal and Academy of Management Review). Then they ranked the sample articles according to simple citations received and citations that were a major basis of the citing study. They suggested that their method provides additional information beyond citations and that the importance of articles that they ranked vis-à-vis the other cited articles examined changed when looking at whether citations were merely in passing or core to the citing article. In fact, using their data listed in Tables 2 and 3 of their study, we calculated the following correlations between rank based on simple citations and rank based on whether the citation was core to the citing article: Spearman $\rho(39)=.71$, $p<.001$ and Spearman $\rho(31)=.66$, $p<.001$ for Academy of Management Journal and Academy of Management Review respectively. Evidently, using simple citations is a good enough proxy for influence, particularly because their method is, as they admit, rather "arduous and time-consuming" (Kacmar & Whitfield, p. 401).

3. Papers = Number of papers published as listed in WoS General Search. Here we only consider number of papers indexed in the WoS database. To the extent that scholars mostly concentrate on publishing in journals, this number is a relatively accurate indicator of total research output. Still, this number is approximate because it excludes book chapters, books, and journals not listed in ISI (but which have been cited in journals listed in ISI). Bibliometricians

who wish to know the exact number of papers should consult the authors's CV to obtain information of total output. Given that obtaining this information might not be feasible, excluding this information will still provide a good enough approximation, particularly because we control for age (see below). In any case, although using number of papers listed in WoS and not total actual papers underestimates output, it is still useful to include this measure, along with total citations received from all sources, in the construction of *IQp*. The more cites one receives from all sources the more one has had an impact on the field (particularly in cases where one has received many citations from publications not listed in ISI's General Search). Individual should get "extra credit" for these latter cites because they did not figure in ISI's General Search but got cited anyway by others thus attesting to the usefulness of the publication (note, peripheral citations are generally more difficult to access; thus, if they are cited, it suggests that the citing author went to lengths to obtain the publication). Researchers who publish in outlets not listed in WoS (e.*g*., book chapters) are taking a risk because their work will not be as widely exposed as work published in WoS-indexed journals. Thus, these individuals should get rewarded for this work that is cited. Individuals whose peripheral work gets cited are certainly more influential per WoS-listed paper than are others whose non-WoS listed work is not cited. Note, self-citation inflation is corrected for in the *IQp*, because for every self-citation a paper must be published (which thus reduces *IQp*). Finally, our formula does include, to a degree, a correction for including citations of all papers (including those not indexed in WoS) by correcting for age, which can also be seen as a proxy for quantity of "peripheral" output.

4. Age = how many years ago the author entered the field (i.e., received the terminal degree). To estimate scientific age one must know when the author received the degree either by consulting the author's CV (an internet search will usually suffice) or simply by finding a reference to the author's thesis (most researchers will cite their thesis--if not, then looking up the reference to the author's thesis in UMI). In all other instances or if the author does not have a doctorate (e.g., like in the case of Y. Chauvin, a Nobel laureate in physics), the date of the first

published paper should be used to determine academic age. Podsakoff received his doctorate in 1980 and is thus "37" years old.

Putting the above information together gives Podsakoff the following *IQp*

$$IQp = \frac{3650}{50 + \frac{37 * 1.046(50+1)}{2*50}}$$

$$= 52.34$$

What is the quality and productivity contribution to this *IQp*? We calculate that Podsakoff would have received 986 (=1.046 * 37 * (50+1) / 2) total citations if his work were of average quality in his top three fields, or an estimated 19.72 citations per paper published (=1.046 * 37 * 51 / 2* 50). In reality, Podsakoff's work is cited 3650 times, or 3.7 times more frequently than the average paper in his field. *IQp* measures productivity by correcting the 50 papers Podsakoff published by the impact per paper in his top three fields. Impact per paper is 0.28 (19.72 est. citations / paper relative to 50 papers + 19.72 est. citations / paper), so Podsakoff's productivity is 14.15 "adjusted papers". This results in an *IQp* of 52.34 consisting of 14.15 adjusted papers cited 3.7 times more frequently than the standard paper in Podsakoff's top three fields. *IQp*

How does a 52.34 compare with other prominent scholars, Nobel winners in exact and social sciences, and more "mortal" scholars? In the following section, we describe how we calibrate *IQp* and how this index compares to *h.*

Testing *IQp*

To see how *IQp* works in practice, and how we can differentiate it from *h*, we initially collected data on ten towering and relatively current individuals in psychology (we refer to this group of Psychologists as "great psychologists"). We used psychology as the initial starting point because it is a field familiar to the first author and because no Nobel prizes are given in this field (thus, there is no reverse causality of sorts that increase citations due to the celebrity emanating from having won a Nobel).

We also selected Psychology as a starting point because it is a social science and similar in method and outlook to Economics (the domain of the second author), which is a social science discipline that does get rewarded with a Nobel--thus, we can examine *IQp's* of officially-recognized scholars who are truly great. We therefore gathered data on the last 10 Nobel prize winners in Economics[vi]. Note, as concerns the great psychologists, we used our professional judgment in selecting these 10 individuals, who in our opinion have made substantial and original contributions to Psychology as a discipline. Assuming that we had correctly selected the 10 psychologists, we expected to find no difference in the *IQp* of the psychologists and economists (or another group of officially-recognized eminent psychologists[vii]). A truly great individual in psychology will be roughly equivalent to a truly great individual in economics.

H5: The mean IQp of a sample of great psychologists and Nobel economists will be approximately equal.

We also collected data on the last 10 Nobel laureates in Physics, Chemistry, and Medicine[viii]. Again, assuming that the above 40 individuals are all truly exceptional we expected to find no difference in the *IQp* between the Psychologists, and the Nobel winners in Economics, Physics, Chemistry and Medicine.

H6: The mean IQp of a sample of great psychologists, Nobel laureates in economics, physics, chemistry, and medicine will be approximately equal.

Finally, to test the index appropriately, it should distinguish prominent individuals from other less well-known individuals. We thus gathered data on 30 randomly selected scholars, 15 of whom were members of the Society of Industrial and Organizational Psychology (SIOP) and who were either assistant or associate professors (to ensure that they were not too academically "old," and thus ensure that we have variance on the variables concerned)[ix]. The remaining 15 scholars were fellows of SIOP (note, being nominated a fellow is not an easy affair at SIOP; these individuals must have made a substantial impact on their field and must be senior figures and should have a minimum academic age of 10 years prior to being nominated[x]). Thus, the impact of Fellows and Members on the field should not be equivalent, as tested in the following hypothesis:

H7: The mean IQp of a sample of SIOP fellows will be significantly higher than that of SIOP members.

However, both SIOP members and fellows should be in a different league as compared to the four groups of highly prominent scientists listed above.

H8: The mean IQp scores of the SIOP fellows and members will be significantly lower than the mean IQp of the four groups of highly prominent scientists.

Furthermore, the *IQp* scores of SIOP members should be significantly different for Assistant Professors and Associate Professors (and both these groups should have lower mean scores than SIOP Fellows). Of course, we are assuming here that SIOP members that have been promoted to associate professor status at their respective universities were promoted because, in part, they had impact on their respective fields of research. However, because assistant professors would probably be more junior (or alternatively they may be older individuals who had not received tenure because they did not have an impact), their research impact should be significantly lower.

H9a: The mean IQp scores of the fellows will be significantly higher than the means scores of members at the rank of assistant professor and of associate professor.

H9b: The mean IQp scores of the members at the rank of associate professor will be significantly higher than the mean IQp of members at the rank of assistant professor.

As a final calibration of the *IQp*, we predicted that peer ratings of scholarly greatness would be strongly related to *IQp* scores. We obtained peers ratings of scholarly greatness for the great Psychologists and Nobel Economists. Domain-level experts, including nine psychologists and eight economists (all from various areas in these disciplines) ranked-ordered the great Psychologists and Nobel Economists, respectively, in order of their greatness. However, as noted previously, *h* underestimates scholarly impact. Thus, we would expect that peer ratings of scholarly greatness would be weakly related with *h* scores.

H10: peer ratings of scholarly greatness will correlate strongly with IQp

H11: peer ratings of scholarly greatness will correlate weakly with h

Results

Refer to Tables 1 and 2 for data and descriptive statistics on the individuals included in the study (note: all data were retrieved in April 2007).

[Insert Table 1 & 2 about here]

Based on the data reported in Table 1, one of the limits of *h* is evident. Kahneman and Eysenck (both Psychologists, although Kahneman did win the Nobel in economics) have about the same *h* (i.e., 52 and 50 respectively). However, Kahneman's 52 most-cited papers listed in ISI have 24450 citations (and on average 470 cites each!) whereas Eysenck's top 50 papers have 6812 cites (and on average only 136 cites each).

Clearly the impact of Kahneman's papers (included in *h*) have been far greater than Eysenck's, even though Kahneman is academically younger than Eysenck (by 21 years). This difference is not captured by *h*. Both of these scholars have had about the same cites but Kahneman generated these cites with about ten times fewer papers. It is thus evident that Kahneman has put out better quality papers, information that *h* does not capture. Also, Asch's *h* is dwarfed by Eysenck's *h*. However, Eysenck received on average 31.91 cites per paper whereas Asch received 242.87 cites per paper. Furthermore, Kahneman has two papers with over 3,000 cites each and five papers with over 1,000 cites each. Eysenck most cited paper has only received 819 cites. Based on these cases it appears that *h* favors quantity over quality. *IQp*, however, appears to set the record straight, giving Kahneman a score of 207.97 and Eysenck only 30.65.

Turning to the correlation matrix for the variables it is evident that *h* correlates strongly with citation counts, $r(80)=.91$, $p < .001$, providing almost the exact same estimate as have recent studies, for example, Spearman $\rho(31) = .90$, $p < .01$ (Cronin & Meho, 2006); $r(55) = .87$, $p < .001$

(Saad, 2006); $r(147) = .94$, $p < .001$ (van Raan, 2006). This finding is troubling because *h* and citations seem to be measuring one and the same thing (and that citations depend heavily on output and discipline, ceteris paribus). Furthermore, *h* is strongly related to $\sqrt{number\ of\ \ papers}$ as well as to *c* (the correction factor). Finally, although *IQp* and *h* are correlated ($r = .47$, $p < .001$) suggesting a degree of convergent validity for our index, the relation is moderate. Note, as we suspected, *h* is unrelated to the quality measure; however, it is strongly related to the productivity measure, $r = .86$, $p < .001$. Furthermore, quality is negatively correlated with productivity ($r = -0.24$, $p < .05$, cf. Bergh et al., 2006). Interestingly, the correlation between *IQp* and age is significantly positive. Note, however, that this association arises due to older scholars being in the highly cited groups of Nobel winners, great psychologists and SIOP fellows. Once we control for these group nestings, this relationship becomes insignificant (refer to Table 4).

[Insert Table 3 about here]

Figure 1 shows the relationship between the quality component of *IQp* and the productivity component of *IQp*. There is a striking separation of disciplines in the quality-productivity space (see also Table 2). Great Psychologists and Economics Nobel prize winners appear to specialize in quality rather than productivity. In contrast, Nobel winners in Medicine, Chemistry and Physics specialize in productivity rather than quality. This contrast is perhaps indicative of the differences in research philosophies between the social sciences and the exact sciences.

[Insert Figure 1 about here]

Turning to the hypotheses now, we estimated the following two-stage least squares (2SLS) regression model using STATA's robust correction on standard errors (given that we detected problems with heteroskedasticity in the ANOVA models, as discussed later):

First-stage: $$\text{Cites} = \alpha_0 + + \beta_2\text{papers}^{\frac{1}{2}} + \sum_{k=2}^{7} \beta_k \text{D}_k + \delta_{ij}$$

Second-stage: $$h = \beta_0 + \beta_8\text{Cites} + \beta_9\text{Age} + \beta_{10}\text{Age}^2 + \sum_{k=11}^{16} \beta_k \text{D}_k + \varepsilon_{ij}$$

D refers to dummy variables indicating group membership (to control for all unobserved domain-specific effects that might predict *h*[xi]). Please refer to Table 4 for estimated parameters (note that adding papers to the model did not increase its predictive validity significantly, i.e., the coefficient for papers was insignificant). Note we mean-centered age so as to avoid collinearity with the quadratic term.

[Insert Table 4 about here]

A large proportion of the variance in citations (75.11%) was predicted by the first-stage model. Furthermore, almost all the variance in *h* (92.97%) was predicted by the second-stage model. As we suspected, the quadratic age term was negative, providing support for *H1*. We plotted the fitted model as a function of age (the partial coefficients). The slope of the line climbed rapidly and then tapered off. Solving for the first derivative of *h* with respect to age, we found that the lines is completely flat (β=.00) when age is 23.27 and *h* is 17.01. This result provides support for *H1*--it becomes increasingly difficult to increase *h* with age (note: the age and age^2 term were not significant when using the same predictive model with *IQp* as the dependent variable).

As suggested in *H2*, *h* would be driven by cites as its proximal predictor and $\sqrt{number\ of\ \ papers}$ as its distal predictor. We have evidence for this causal model from two methods. First, the 2SLS regression results show that cites is a significant predictor of *h*, and that cites depends largely on $\sqrt{number\ of\ \ papers}$. Second, a Sobel test indicated that the effect of $\sqrt{number\ of\ \ papers}$ on *h* was fully mediated by cites, $z = 3.58$, $p < .01$. When the mediator (cites) was excluded from the model, the direct effect was significant and large (standardized β = .64). When including the mediator (cites), the standardized indirect effect of

$\sqrt{number\ of\ \ papers}$ was β = .67, whereas the direct effect became insignificant and small (standardized β = .18).

It appears evident therefore that *h* depends largely on quantity of output, which can be viewed as its distal predictor. This causal pattern was not found with *IQp* (see Table 4). That is, cites did not predict *IQp*, which is obvious given how we constructed the *IQp*. These results provide support for *H2*.

To demonstrate that *h* cannot be useful for cross-disciplinary comparisons, we estimated a one-way ANOVA model using the Brown-Forsythe robust test (given that Levene's test was violated). Although the *F*-test indicated the means of *h* differed significantly as a function of group, $F(6,27.03)=16.79$, $p < .001$, the Games-Howell post-hoc tests indicated the groups differed in ways that were not theoretically defensible, as indicated in Table 5.

[Insert Table 5 about here]

For instance, the great Psychologists did not have a significantly higher *h* than did the SIOP Fellows; the Nobel economists were significantly lower than the Nobel winners in Medicine and were no different from the SIOP Fellows. These results provide support for *H3.*

Turning to the performance of the *IQp*, we first examined how much our correction factor *c* would attenuate *h* directly (by using it to "deflate" *h* for field-specific average citation rates--we achieved this correction by dividing *h* by *c*). The reduction in *h* was for physicists 59.57% (i.e., the mean *h* was reduced from 45.1 to 18.23) and for chemists it was 68.92% (i.e., the mean *h* was reduced from 62.4 to 19.39). Interestingly, our degree of correction was remarkably similar to the degree of correction estimated by Batista et al. (2006) (who corrected *h* using an entirely different approach, i.e., for the number of coauthors). Their correction reduced the *h* of physicists by 53.4% (i.e., the mean *h* was reduced from 19 to 8.86) and the mean *h* of chemists by 67.51%. Thus, our correction factor produces a very similar correction outcome result in a simpler and more efficient manner.

Turning to *H4,* we estimated an ordered probit model (with robust standard errors) using the components of *IQp* and *IQp* itself (i.e. the two main effects of quality and productivity and their interaction). This model tests whether greatness--ordered from (a) great psychologists and Nobel winners, to (b) SIOP fellows, and (c) SIOP members--is associated with quantity, productivity, or the interaction of the two. Results indicate that the main effects of quality (β = -0.96, $p > 0.10$), productivity (β = .58, $p > 0.10$) were not significantly related to academic achievement, whereas their interaction, that is, the *IQp* (β = 0.49, $p < 0.05$) was. The overall predictive power of the model was very high ($Pseudo\text{-}r^2$ = .93) and produced suspect standard errors. We thus re-estimated the same model using a generalized ordered model (Williams, 2006) and found essentially the same result: the effect of *IQp* was significant. This result suggests that the interaction of quality and productivity (i.e., *IQp*) are the drivers of scholarly achievement, providing support for *H4*.

Next, we estimated a one-way ANOVA model, again using the Brown-Forsythe robust test (because Levene's test was violated) to examine whether the mean *IQp's* differed significantly as a function of group, which they did: $F(6,27.21)=14.23$, $p < .001$. This time, the Games-Howell post-hoc tests indicated the groups differed in ways that we had predicted, as indicated in Table 6. These results provide support for *H5, H6, H7, and H8.*

We then examined how the *IQp* performed with the SIOP fellows and members only, differentiating the sample by professorial rank. Results indicated that the means of the three groups differed significantly, Brown-Forsythe robust $F(2,19.33)=19.51$, $p < .001$. Games-Howell post-hoc tests indicated that assistant professors had significantly lower *IQp's* than did the associate professors and fellows ($p < .05$ and $p < .001$ respectively). Moreover, associate professors had lower *IQp's* than the fellows ($p < .01$). This result supports *H9a* and *H9b*.

Finally, we examined the extent to which *IQp* and *h* correlated with peer ratings of scholarly greatness (of the great psychologists and economist group). Expert raters, that is eight professors in economics and nine in psychology rank-ordered from highest to lowest (in terms of greatness) the individuals in the group of great psychologists or the Nobel economists respectively. Rank order went from 1 to 10 (thus, a lower average rank indicated higher greatness). Average rankings received are included in the last column of Table 1.

To ensure that there was some measure of homogeneity in the rankings (thus justifying an aggregation of the scores), we computed the Cronbach α reliability coefficient (where raters were the "items" and ratees the *n*-size). The α was .70 for the Economics group and .81 for the Psychology group, both of which are in acceptable ranges. We also computed James, Demaree, and Wolf's (1984) r_{wg} index of rater agreement to justify aggregation. We used the recommendations of Lindell and Brandt (1999; see also Lindell, Brandt, & Whitney, 1999) and assumed a maximum variance distribution (with a Spearman-Brown correction). The results indicated that the average $r_{wg(j)}$ index for the great Psychologists .74, whereas it was .64 for the Economists (with the latter being a bit below conventionally required levels). Evidently, the ratings of the psychologists were quite cohesive and reliable, and those of the economists not as reliable as we might have hoped.

We first pooled the data of the two groups and correlated average rankings with the *h* and *IQp* scores. For *h*, the correlation was in the right direction $r(20) = -.39$, $p < .10$ and modestly strong; we report a higher *p*-value here--at the .10 level--given that the small sample sizes. For *IQp*, the correlation was also in the right direction $r(20) = -.59$, $p < .01$ and stronger. The difference in the dependent correlations (i.e., between r_{xy} and r_{vy}, see Cohen & Cohen, 1983) was not significant, which is not surprising given the small sample size.

Next, we examined each group individually. The *h* index had a strong correlation with the peer ratings for the economist group, Spearman $\rho(10) = -.77$, $p < .01$. The *IQp* correlated less strongly, $\rho(10) = -.56$, $p < .10$ (the difference in dependent correlations was not significant). Finally, for the psychologist group (which had the most reliable rankings), *h* did not correlate with the peer rankings, Spearman $\rho(10) = -.09$, $p > .10$. The *IQp* correlated strongly, $\rho(10) = -.80$, *p* < .05 and surprisingly for the small sample size, the differences in the dependent correlations was significant, *t*(7)=3.29, *p* < .001, two tailed.

Overall, these results provide support for *H10*. The results also tentatively suggest that the *IQp* predicts peer ratings better than does the *h* index. As concerns *H11,* the results were generally supportive, given that *h* was uncorrelated with peer rankings of the psychologists (and these peer ratings were quite reliable). Although it correlated well with the economist ratings, the overall correlation with the pooled group was quite weak.

Discussion

The results of this study suggest that the *h* index does have important limitations that question its usefulness as an indicator of scholarly quality. Even though newer versions of *h*, which are supposed to improve *h*'s precision, have recently been proposed (e.g., Egghe, 2006; Jin, 2007), these indexes suffer from the same limitations that *h* does and do not necessarily identify unique information on scholarly output or quality[xii]. What *h* seems to tap is quantity more than anything else. As such, this index should not be used to separate the wheat from the chaff.

The *IQp* did show that it was useful both for within disciplines and across disciplines. Our index is, of course, limited. First, it is only as useful as the data on which it is based (and we relied entirely on the Thompson databases in this regard). Although this limitation is apparent our index is not defined exclusively for use with Thompson data. As with *h, IQp* can be constructed based on other databases (e.g., Scopus and Google Scholar, as long as they have all the information necessary for constructing the *IQp*. To our knowledge, only Thompson has information on field-specific journal impact factors). Second, the *IQp* does not account for the time dependence of citations to publications. The *IQp* quality indicator will be larger for researchers whose papers have been published earlier compared to researchers who have published their papers more recently ceteris paribus (i.e., for researchers of same age, same number of papers published). Thus, the *IQp* favors early publications to more recent publications.

From a practical perspective, and based on the data we have gathered, the *IQp* could be useful for hiring and promotion. Institutions should, of course, set their own criteria based on what they deem to be reasonable benchmarks and also take into account the quality of research papers published by the person being evaluated, particularly if the target individual is young in terms of academic age. If impact is an important criterion, then based on our sample data (using the Huber ML estimates as a basis) it seems reasonable to expect that promotion to tenure (Associate Professor) would require a *IQp* of about 5. Promotion to professor should require an *IQp* of about 8-10. We would expect a chaired professor to have an *IQp* of over 10, at a minimum. A *IQp* of over 15 would be indicative of a scholar who has had a rather substantial impact on the field. Anything over 20 would indicate that the scholar has had a very important influence on the field.

So, what is indicative of greatness? An *IQp* of about 40 should do the trick. If one is fortunate enough to be in a discipline that is rewarded by a Nobel, a *IQp* of over 60 should easily put one in Nobel Prize territory (only about a third of Nobel winners had a *IQp* below 60). An *IQp* higher than 80 would be indicative of a towering figure in any field (the mean *IQp* of the great psychologist group was 89.15!). An *IQp* of over 100 would be rare, indicating true colossus status.

Prior to concluding, we looked beyond this sample to see how the *IQp* quantifies the impact of other prominent individuals. Ed Witten (cited in Ball, 2005), who is highly cited (and has an *h* of 115) is considered as the most "brilliant living physicist" (Ball, 2005, p. 900). Witten's *IQp* is 230 *IQp*--and certainly it is much higher than Greengard's *IQp* of 68, a Nobel winner in medicine included in our sample who has an *h* of stratospheric proportions (i.e., 145). The highest Nobel *IQp* we calculated was for Kahneman (207), though his long time colleague, the late Amos Tversky has a *IQp* of 225! Some top biologists cited in Hirsch (2005) like SH Snyder and RC Gallo have *h's* of astronomical proportions: 198 and 155 respectively! However, they do not look nearly as impressive with *IQp's* of 117 and 75 respectfully (although these are truly remarkable scholars with very high *IQp's*).

Conclusion

It is an unfortunate fact that the majority of published papers never get cited (Hamilton, 1990; Meho, 2007). Publishing pressure may be, to an extent, making scholars produce more papers and not necessarily better quality papers. The consequence of putting quantity ahead of quality instead of focusing on both aspects (or more on quality) might be at a cost to society in direct ways (e.g. the salaries of the scholars, indexing and printing costs, storage costs, and so forth) as well as indirectly (i.e., doing research that might not be useful to society). Would society not be better off if scholars were rewarded for producing fewer, but better quality papers (ceteris paribus, *IQp*'s would stay similar as output goes down but quality goes up; however, fewer societal resources would be used to generate those *IQp*'s)?

Indeed, impact of an article should reflect its quality (in terms of methodological rigor among other factors), which is precisely what Bergh et al. (2006) recently found. They also found that brute quantity was not necessarily rated to impact and that authors with a lower number of highly-cited articles had the most impact. Based on their findings, Bergh et al. noted that a "high-volume article strategy required by many prestigious research institutions for promotion and tenure is unlikely to produce high-impact articles unless those articles also have high citation counts, which is very rare" (p. 97).

To conclude, the results of our study demonstrate clearly that Hirsch's *h* simply does not hit the spot. It seems to measure brute force of output but does not quite capture the finesse of input. We trust that researchers will develop alternative measures that better serve the academic milieu by capturing both dimensions of scholarship that matter; quality and quantity.

References

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

Bergh, D. D., Perry, J., & Hanke, R. (2006). Some predictors of SMJ article impact. *Strategic Management Journal, 27*, 81-100.

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

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.

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

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral science*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Egghe, L. (2006). An informetric model for the Hirsch-index. *Scientometrics, 69*, 121-129.

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

Glänzel, W. (2006b). On the opportunities and limitations of the h-index. Science Focus, 1, 10-11.

Edwards, J. R., & Bagozzi, R. P. (2000). Psychological Methods, 5, 155-174.

Hamilton, D. P. (1990). Publishing by—and for?—the numbers. *Science, 250,* 1331-1332.

Hamilton, D. P. (1991). Research papers: Who's uncited now? *Science, 251,* 25.

Haggbloom, S. J. et al. (2002). The 100 most eminent psychologists of the 20th century. *Review of General Psychology, 6*, 139-152.

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

James, L. R., Demaree, R. G., & Wolf, G. 1984. Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69,* 85-98.

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

Kacmar, K. M., & Whitfield, J. M. (2000). An additional rating method for journal articles in the field of management. *Organizational Research Methods, 3*, 392-406.

Lindell, M. K., & Brandt, C. J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of the CVI, T, r_{wgj}, and r^*_{wgj} indexes. *Journal of Applied Psychology, 84,* 640-647.

Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A revised index of interrater agreement for multi-item ratings of a single target. *Applied Psychological Measurement, 23,* 127-135.

Meho, L. We. (2007). The rise and rise of citation analysis. *Physics World, 20,* 32-36.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46,* 323-351.

Oppenheim, C. (2007). Using the *h*-index to rank influential British researchers in information science and librarianship. *Journal of the American Society for Information Science and Technology, 58*, 297-301.

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.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables." *The Stata Journal* 6, 58-82.

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

Notes:

[i] The denominator of this term is a composite or formative measure (see Edwards & Bagozzi, 2000). Formative measures are usually composed of dimensions that might be independent (a good example of a composite measure is socio-economic status--composed of education, income, and social status--and the UN's Human Development Index--composed of life expectancy, education, and standard of living). Our composite measure serves as a benchmark that requires at least one citation per paper published and at least the standard rate of citation expected for a scholar's age. Composite measures are statistically constructed using principal components analysis (a procedure related to factor analysis). Using our dataset, we extracted the first principal component, which explained 74.89% of the variance in the construct. The loadings on the construct were .87 (i.e., the correlation of the factor score coefficient with each dimension was .87. If the dimensions were orthogonal the expected correlation would be only .71). However, when calculating an *IQp*'s for an individual it is not possible to extract the first principal component (i.e., with an n=1). Thus, for practical purposes, we determined whether it is empirically justifiable to construct the composite by adding the two dimensions. The correlation of the composite measure with the factor score was close to unity (i.e., $r = .92$, $p < .001$) and the average correlation between the items and the composite was $r = .80$, $p < .001$ (which is a bit lower than the .87 of the composite to the items). The reason why the correlation of the items with the composite is rather high is because the two items are themselves significantly correlated, $r = .50$, $p < .001$. Thus, this approximation (i.e., the addition of the two dimensions) is empirically justified.

[ii] We assume that the author published p papers in his or her scientific age a in a field with average citations per year of c. Thus, the first paper will receive a total of $c*a$ citations. The second paper is published after a/p periods and can therefore be cited for $(p-1)/p * a$ periods totaling $c * a * (p-1)/p$ citations. The second paper will be cited $c * a * (p-2)/p$ times, and so forth. Thus, the total number of citations for the author's work is $c * a / p * (1 + 2 + 3 + \ldots + p)$ or $c * a/p * (p+1)*p/2 = c*a * (p+1)/2$ times.

[iii] Batista et al. (2006) suggested making the *h* index comparable across disciplines by controlling for number of authors in the considered *h* papers. Although this approach is useful, it is very labor intensive. Our correction, if applied to *h*, produces a very similar result (refer to the results section).

[iv] The *c*'s we report for our sample of individuals are based on Journal Citations Report data of 2005 (the report that was available to us in April 2007). Given that these aggregates might change from year to year, both within and across fields, we analyzed the data for the 52 fields we used, from 2003 to 2006 (the data for 2006 were released in summer 2007 by SSCI, just before we began working on revised draft of our paper). We analyzed the aggregate impact factors for the 52 subject categories (list in the footnote of Table 1) using a random effects model. The intraclass correlation coefficient was .98. This result indicates that the measures are highly stable within category. However, the aggregate impact factors across the 52 subject categories did increase slightly between 2003-2006, $\beta = .08$, $p < .001$ (from a mean of 1.93 to 2.17; the *SD's* for all years were almost the same, at 1.20, rounded). Thus, when calculating *SQi*'s, current *c*'s from the same year must be used.

[v] Indeed, using a sample of individuals from our database indicated that the majority of an individual's citations were from the first three subject categories, which follows the typical power-law distribution, Newman (2005).

[vi] We excluded the following individuals who have common names and initials (as well as because of inconsistencies with their initials): McFadden, Prescott, Smith, and Spence (Economics).

[vii] Eight of the ten individuals we selected were included in the Haggbloom et al. (2002) list of the 100 most eminent psychologists (interestingly, the two individuals who were not on this list were given the lowest rating by our expert group, as discussed later). However, because our sample of ten psychologists was not randomly selected, we compared it to a random sample from the top-30 of the Haggbloom et al. list. This random sample of psychologists, from which we excluded individuals on our original list of ten, was not significantly different from the original sample of psychologist on any of the measured criteria, including *h* (28.30 vs. 34.90) and *IQp* (100.35 vs. 105.23), as well as on the other measured variables (i.e., academic age, cites received, and papers published). We thus retained our initial sample given that it was more current. Note that the random sample was also not significantly different from the other eminent individuals we examined (i.e., the Nobel winners); however, they were significantly different from the non-eminent individuals (in the same way the original psychologist sample is, as reported in Table 5). Thus, our original sample is representative of great psychologists. The random sample of eminent psychologists included:

Allport, GW; Bandura, A; Brunner, JS; Cattell, RB; Hebb, DO; Guilford, JP; McClelland, DC; Miller, NE; Rogers, CR; Thorndike, EL.

[viii] For the reason noted previously, we excluded the following: Hall, Mather, Davis, Koshiba, Giacconi, and Gross (Physics), MacKinnon and Tanaka (Chemistry), and Marshall, Warren, Mansfield, Brenner Hunt, Nurse, and Carlsson (Medicine).

[ix] We excluded individuals who had common names or inconsistencies with initials.

[x] The criteria for nomination as a SIOP fellow can be viewed at the following web-page: http://www.siop.org/Fellows/siop_fellowship.aspx

[xi] The great psychologists group served as the reference group, followed by Economics, SIOP Fellows, Physics, SIOP Members, Chemistry, and Medicine.

[xii] For example, the Egghe *g*-index is largely redundant. Based on the data reported in Table 2 by Egghe (p. 143), we calculated the correlation between h and g to be Spearman $\rho(14) = .97$, $p < .001$.

Table 1: Data of scholars in various groups

Name	Group	Field code*	*c*	Cites	Papers	Age	*h*	Qual.	Prod.	*IQp*	Peer score
ASCH, SE	Psychology	47, 41, 46	1.70	5586	23	75	14	3.66	17.08	62.53	5.00
CRONBACH, LJ	Psychology	44, 46, 17	1.16	23572	84	67	23	7.14	26.77	191.18	4.78
DAWES, RM	Psychology	47, 46, 24	1.36	6335	105	44	28	2.00	23.44	46.86	7.33
EAGLY, AH	Psychology	47, 46, 42	1.37	8280	104	42	36	2.74	22.71	62.23	7.89
EYSENCK, HJ	Psychology	47, 46, 43	1.51	39246	1230	67	50	0.63	48.50	30.65	6.67
FESTINGER, L	Psychology	47, 46, 41	1.56	15457	47	65	24	6.34	24.66	156.33	4.11
MASLOW, AH	Psychology	46, 41, 17	1.58	9673	66	73	17	2.51	31.01	77.69	7.33
SKINNER, BF	Psychology	46, 45, 2	1.86	24011	159	76	33	2.13	49.10	104.38	3.22
TULVING, E	Psychology	45, 28, 41	2.58	24092	187	50	65	1.98	48.20	95.63	6.33
TVERSKY, A	Psychology	16, 47, 24	1.05	31537	118	42	59	12.00	18.74	224.83	2.33
AKERLOF, GA	Economics	16, 23, 6	0.95	4757	52	41	22	4.63	14.32	66.28	4.75
AUMANN, RJ	Economics	16, 50, 25	1.00	4570	51	52	18	3.38	17.46	58.93	7.13
ENGLE, RF	Economics	16, 50, 52	0.96	11977	79	38	30	8.20	14.98	122.85	6.13
GRANGER, CWJ	Economics	16, 50, 52	0.96	15359	164	48	34	4.04	20.33	82.04	4.88
HECKMAN, JJ	Economics	16, 50, 21	0.88	10607	106	36	34	6.29	13.83	87.01	3.51
KAHNEMAN, D	Economics	45, 16, 41	1.68	36153	135	46	52	6.89	30.16	207.97	4.25
KYDLAND, FE	Economics	16, 6, 50	0.82	2832	21	34	10	9.27	8.59	79.67	6.00
PHELPS, ES	Economics	16, 6, 50	0.82	3808	71	48	18	2.70	15.53	41.90	6.25
SCHELLING, TC	Economics	16, 22, 40	0.75	5956	92	56	13	3.05	17.23	52.61	8.88
STIGLITZ, JE	Economics	16, 6, 39	0.80	14924	197	40	50	4.73	14.82	70.06	3.25
ABRIKOSOV, AA	Physics	34, 35, 32	2.04	15606	204	52	30	1.44	42.25	60.66	-

CORNELL, EA	Physics	30, 33, 35	2.10	9229	81	17	37	6.32	14.75	93.19	-
GINZBURG, VL	Physics	1, 35, 32	3.14	21775	459	67	42	0.45	85.83	38.57	-
GLAUBER, RJ	Physics	35, 36, 37	2.37	17253	72	58	35	3.44	35.39	121.84	-
HANSCH, TW	Physics	30, 33, 35	2.10	15046	364	38	63	1.04	35.99	37.25	-
KETTERLE, W	Physics	30, 33, 35	2.10	13017	150	21	50	3.92	19.30	75.61	-
LEGGETT, AJ	Physics	35, 34, 33	2.26	14141	143	43	41	2.03	36.40	73.72	-
POLITZER, HD	Physics	37, 35, 1	3.01	9198	62	33	35	2.94	27.82	81.78	-
SMOOT, GF	Physics	37, 1, 35	3.24	18428	115	37	45	2.65	39.66	104.98	-
WILCZEK, F	Physics	37, 35, 1	3.01	23270	322	33	73	1.45	43.14	62.58	-
AGRE, P	Chemistry	3, 8, 38	4.55	14747	294	33	64	0.67	59.98	39.93	-
CHAUVIN, Y	Chemistry	11, 12, 10	2.64	3303	114	43	26	0.51	38.07	19.30	-
CIECHANOVER, A	Chemistry	3, 8, 4	4.54	18132	164	26	62	1.86	43.63	81.15	-
FENN, JB	Chemistry	9, 51, 13	2.05	9467	97	67	33	1.41	40.43	56.92	-
GRUBBS, RH	Chemistry	12, 11, 10	2.54	32104	606	39	88	1.07	45.85	48.97	-
KNOWLES WS	Chemistry	12, 11, 10	2.54	3057	38	65	19	0.95	26.23	24.91	-
KORNBERG, RD	Chemistry	3, 8, 4	4.54	21549	218	35	77	1.24	58.47	72.34	-
ROSE, IA	Chemistry	3, 4, 8	4.13	9867	194	55	56	0.45	71.84	32.03	-
SCHROCK, RR	Chemistry	10, 12, 11	2.32	27434	573	36	86	1.15	38.97	44.62	-
WUTHRICH, K	Chemistry	3, 4, 11	3.70	67559	673	43	113	1.26	71.23	89.76	-
BLOBEL, G	Medicine	3, 8, 4	4.54	46373	407	40	120	1.25	74.43	93.10	-
FIRE, AZ	Medicine	3, 8, 20	4.78	12362	86	24	46	2.48	34.65	85.83	-
FURCHGOTT, RF	Medicine	31, 38, 7	2.93	29650	146	67	54	2.05	58.94	121.10	-
GREENGARD, P	Medicine	3, 28, 31	3.80	69884	921	54	145	0.74	92.50	68.26	-

HARTWELL, LH	Medicine	3, 8, 20	4.78	19579	135	43	66	1.40	58.60	82.08	-
HORVITZ, HR	Medicine	3, 8, 28	4.63	28728	209	33	88	1.79	56.11	100.55	-
KANDEL, ER	Medicine	28, 3, 38	3.75	50642	477	51	121	1.11	79.74	88.42	-
LAUTERBUR, PC	Medicine	11, 49, 12	2.67	7081	136	45	42	0.86	41.88	36.03	-
MELLO, CC	Medicine	3, 8, 15	4.89	8782	103	17	33	2.03	29.81	60.59	-
SULSTON, JE	Medicine	3, 20, 8	4.60	22170	106	41	50	2.20	50.13	110.23	-
AGUINIS, H	SIOP F.	42, 24, 47	1.13	625	49	14	13	1.58	6.95	10.95	-
BURKE, WW	SIOP F.	24, 42, 5	1.05	530	41	44	8	0.55	14.97	8.21	-
CLEVELAND, JN	SIOP F.	42, 24, 47	1.13	590	31	25	12	1.30	9.94	12.93	-
DUNNETTE, MD	SIOP F.	42, 24, 46	1.15	3164	87	53	25	1.18	22.72	26.87	-
GREENHAUS, JH	SIOP F.	42, 24, 5	1.06	2267	54	37	21	2.11	14.55	30.67	-
HOFMANN, DA	SIOP F.	42, 24, 5	1.06	706	19	15	11	4.45	5.80	25.82	-
LANCE, CE	SIOP F.	42, 24, 5	1.06	547	42	22	15	1.09	9.28	10.15	-
MARTOCCHIO, JJ	SIOP F.	42, 24, 5	1.06	899	32	20	17	2.58	8.13	20.95	-
MOBLEY, WH	SIOP F.	42, 24, 5	1.06	2008	24	36	11	4.22	10.86	45.82	-
OFFERMANN, LR	SIOP F.	42, 24, 47	1.13	330	22	26	8	0.97	9.06	8.82	-
PERREWE, PL	SIOP F.	42, 24, 5	1.06	479	51	22	12	0.79	9.62	7.62	-
REILLY, RR	SIOP F.	42, 24, 5	1.06	933	54	38	15	0.84	14.84	12.53	-
SPECTOR, PE	SIOP F.	42, 24, 5	1.06	3900	103	32	27	2.22	14.65	32.48	-
STONE, DL	SIOP F.	42, 24, 5	1.06	415	29	26	8	1.01	9.54	9.60	-
THORNTON, GC	SIOP F.	42, 24, 5	1.06	1034	55	41	14	0.85	15.75	13.42	-
ARONSON, ZH	SIOP M.	18, 19, 5	0.89	15	5	9	2	0.63	2.45	1.53	-
COSTANZA, DP	SIOP M.	42, 44, 46	1.19	75	8	11	5	1.28	3.83	4.89	-

DRUSKAT, VU	SIOP M.	24, 42, 5	1.05	139	10	11	7	2.20	3.88	8.51	-
ELICKER, JD	SIOP M.	5, 24, 42	1.00	0	2	7	0	-	1.45	-	-
HOYT, CL	SIOP M.	14, 42, 47	0.97	38	5	4	4	3.25	1.59	5.18	-
JAGACINSKI, CM	SIOP M.	44, 47, 42	1.23	349	22	29	8	0.85	10.09	8.59	-
KARREN, RJ	SIOP M.	42, 24, 47	1.13	238	7	29	4	1.81	5.10	9.23	-
LAHUIS, DM	SIOP M.	42	1.11	7	5	5	1	0.42	2.00	0.84	-
LAPIERRE, LM	SIOP M.	42, 24	1.08	12	6	6	2	0.53	2.32	1.23	-
NAIDOO, L	SIOP M.	47	1.39	2	2	2	1	0.48	1.02	0.49	-
PICCOLO, RF	SIOP M.	42, 24, 5	1.06	46	5	2	3	7.25	1.01	7.34	-
SEIJTS, GH	SIOP M.	42, 48, 24	1.43	155	22	9	7	1.05	5.16	5.39	-
TONIDANDEL, S	SIOP M.	42, 26, 27	1.95	22	13	6	3	0.27	4.25	1.14	-
WESTABY, JD	SIOP M.	48, 42, 46	1.67	52	12	12	4	0.40	5.70	2.27	-
WRZESNIEWSKI, A	SIOP M.	2, 24, 29	2.09	234	10	8	7	2.54	4.79	12.19	-

Notes: SIOP F. = SIOP Fellows; SIOP M. = SIOP Members; Qual.= Quality; Prod. = Productivity; SIOP Members that are italicized are Assistant Professors (and in one case a Senior Lecturer);

* Field codes are: 1=Astronomy & Astrophysics=3.97; 2=Behavioral Science=2.64; 3=Biochemistry & Molecular Biology=4.37; 4=Biophysics=3.05; 5=Business=0.93; 6=Business, Finance=0.74; 7=Cardiac & Cardiovascular Systems=3.60; 8=Cell Biology=5.55; 9=Chemistry, Analytical=2.21 10=Chemistry, Inorganic & Nuclear=2.04; 11=Chemistry, Multidisciplinary=2.99; 12=Chemistry, Organic=2.41; 13=Chemistry, Physical=2.32; 14=Communication=0.75; 15=Developmental Biology=5.10; 16=Economics=0.83; 17=Education & Educational Research=0.53; 18=Education, Scientific Disciplines=0.89; 19=Engineering, Multidisciplinary=0.87; 20=Genetics & Heredity=4.46; 21=Industrial Relations & Labor=0.88; 22=International Relations=0.72; 23=Law=1.23; 24=Management=1.04; 25=Mathematics, Interdisciplinary Applications=1.64; 26=Medical Informatics=1.39; 27=Medicine, Research & Experimental=3.44; 28=Neurosciences=3.55; 29=Nutrition & Dietetics=2.53; 30=Optics=1.79; 31=Pharmacology & Pharmacy=2.60; 32=Physics, Applied; 33=Physics, Atomic, Molecular & Chemical=2.33; 34=Physics, Condensed Matter=1.76; 35=Physics, Multidisciplinary=2.56; 36=Physics, Nuclear=1.76; 37=Physics, Particles and Fields=2.99; 38=Physiology=3.09; 39=Planning & Development=0.83; 40=Political Science=0.57; 41=Psychology=2.27; 42=Psychology, Applied=1.11; 43=Psychology, Clinical=1.93; 44=Psychology, Educational=1.16; 45=Psychology, Experimental=2.04; 46=Psychology, Multidisciplinary=1.47; 47=Psychology, Social=1.39; 48=Public, Environmental & Occupational Health=2.11; 49=Radiology, Nuclear Medicine & Medical Imaging=2.33; 50=Social Sciences, Mathematical Methods=0.94; 51=Spectroscopy=1.66; 52=Statistics and Probability=1.39.

Table 2: Means and Standard Deviations of Various Scholar Groups

	c	Cites	Papers	Age	Quality	Product-ivity	*h*	*IQp*	*IQp*[1]
					Great Psychologists				
Mean	1.57	18778.90	212.30	60.10	4.11	31.02	34.90	105.23	89.15
SD	*0.43*	*11514.88*	*360.92*	*14.05*	3.43	12.73	*17.59*	64.80	
					Economics Nobels				
Mean	0.96	11094.30	96.80	43.90	5.32	16.73	28.10	86.93	74.01
SD	*0.27*	*9957.10*	*54.73*	*7.22*	2.25	5.60	*14.61*	47.98	
					Physics Nobels				
Mean	2.54	15696.30	197.20	39.90	2.57	38.05	45.10	75.01	74.59
SD	*0.50*	*4693.73*	*137.99*	*15.66*	1.71	19.27	*13.53*	27.04	
					Chemistry Nobels				
Mean	3.35	20721.90	297.10	44.20	1.06	49.47	62.40	50.99	49.14
SD	*1.03*	*19065.32*	*232.89*	*13.77*	0.43	15.18	*30.00*	23.85	
					Medicine Nobels				
Mean	4.14	29525.10	272.60	41.50	1.59	57.68	76.50	84.62	86.45
SD	*0.81*	*20387.56*	*264.42*	*14.50*	0.60	20.00	*39.48*	24.81	
					All Nobels				
Mean	2.74	19259.40	215.93	42.38	2.63	40.48	53.03	74.39	71.85
SD	*1.37*	*15990.15*	*199.95*	*12.81*	2.18	21.94	*31.59*	34.36	
					SIOP members				
Mean	1.28	92.27	8.93	10.00	1.64	3.64	3.87	4.58	4.39
SD	*0.36*	*107.50*	*6.22*	*8.30*	1.86	2.40	*2.50*	3.86	
					Asst. Profs.				
Mean	1.18	17.75	5.38	5.13	1.60	2.01	2.00	2.22	1.18
SD	*0.35*	*16.67*	*3.42*	*2.42*	2.50	1.05	*1.31*	2.60	
					Assoc. Profs.				
Mean	1.40	177.43	13.00	15.57	1.45	5.51	6.00	7.30	7.35
SD	*0.37*	*103.69*	*6.35*	*9.27*	0.77	2.13	*1.63*	3.30	
					SIOP Fellows				
Mean	1.08	1228.47	46.20	30.07	1.72	11.78	14.47	18.46	14.78
SD	*0.04*	*1096.46*	23.47	11.22	1.21	4.40	5.87	11.45	

[1]These mean estimates are based on Huber's robust maximum likelihood estimates (which weighs extreme values less). Note, the mean Great Psychologists *IQp*, when not including Tversky, a clear outlier, is 91.94.

Table 3: Correlations among variables for sample

	Cites	Papers	Papers^{½}	Age	*c*	Quality	Productivity	*h*	*IQp*
Cites	-								
Papers	.80***	-							
Papers^{½}	.84***	.95***	-						
Age	.45***	.36**	.49***	-					
c	.55***	.38***	.51***	.16	-				
Quality	.04	-.21	-.16	.09	-.29**	-			
Productivity	.81***	.69***	.82***	.58***	.79***	-.24*	-		
h	.91***	.74***	.86***	.36**	.70***	-.07	.86***	-	
IQp	.59***	.16	.32**	.52***	.28*	.65***	.41***	.47***	-

Note: $n = 80$; ***$p < .001$; **$p < .01$; *$p < .05$

Table 4: Estimated parameters for 2SLS regression model predicting *h* and *IQp*

	Coefficient	Robust Std. Error	*t*	*p*	Coeff.	Robust Std. Error	*t*	*p*
First stage			Dependent variable: cites					
Intercept	-2618.01	3855.89	-0.68	0.50				
$Papers^{1/2}$	1720.99	179.23	9.60	0.00				
Age	31.75	72.50	0.44	0.66				
Age^2	-0.27	2.84	-0.09	0.93				
Group 2	-2656.77	3809.38	-0.70	0.49				
Group 3	-7135.56	3906.29	-1.83	0.07				
Group 4	-4496.11	3731.34	-1.20	0.23				
Group 5	-921.45	5050.44	-0.18	0.86				
Group 6	-4231.79	3702.86	-1.14	0.26				
Group 7	6023.88	3741.07	1.61	0.11				
Adj. r^2	75.11%							
F(9,70)	24.49**							
Second stage	Dependent variable: *h*				Dependent variable *IQp*			
Intercept	15.02	4.71	3.19	0.00	127.18	33.90	3.75	0.00
Cites	0.00	0.00	8.79	0.00	0.00	0.00	-1.07	0.29
Age	-0.31	0.07	-4.18	0.00	-0.07	0.35	-0.21	0.84
Age^2	-0.01	0.00	-4.51	0.00	-0.01	0.01	-0.91	0.36
Group 2	-4.38	4.67	-0.94	0.35	-31.59	35.05	-0.90	0.37
Group 3	-3.21	4.92	-0.65	0.52	-105.50	33.11	-3.19	0.00
Group 4	5.38	4.58	1.18	0.24	-38.39	29.92	-1.28	0.20
Group 5	-8.90	5.09	-1.75	0.09	-111.00	29.44	-3.77	0.00
Group 6	14.89	5.40	2.76	0.01	-59.16	29.29	-2.02	0.05
Group 7	13.05	5.72	2.28	0.03	-19.59	30.80	-0.64	0.53
r^2	92.97%				48.91%			
F(9,70)	109.00**				24.28**			

Note: Note: for second stage regression in model predicting *h*, precise estimates (and standard errors) are: cites = .0017254 (.0001963), age = -.3059197 (.0732689), age^2 = -.0117629 (.0026069). Although the plotted fitted function is as we hypothesized, readers may have noticed that both the main and squared effects of age were negative. The reason for the negative association between *h* and age is because group membership is strongly related to age (Nobels and SIOP fellows being older than SIOP members). A simple regression of *h* and age therefore confounds the effects of group membership with the effects of age. Regressing *h* on age and age^2 the signs on the regression coefficients are positive and negative respectively.

**p < .001.

Table 5: Post-hoc multiple comparisons for *h*

Reference group	Comparison group	Mean diff.	Std. error	Sig.	Mean diff	Std. error	Sig.
			h			*IQp*	
Psychologists	SIOP Fellows				86.79	20.71	0.02
Psychologists	SIOP Members	31.03	5.60	.00	100.64	20.53	0.01
Economists	SIOP Fellows				68.45	15.45	0.02
Economics	SIOP Members	24.23	4.66	.01	82.31	15.20	0.00
Economics	Medicine	-48.40	13.31	.04			
Medicine	SIOP Fellows	62.03	12.58	.01	66.17	8.38	0.00
Medicine	SIOP Members	72.63	12.50	.00	80.03	7.91	0.00
Physicists	SIOP Fellows	30.63	4.54	.00	56.56	9.05	0.00
Physicists	SIOP Members	41.23	4.33	.00	70.42	8.61	0.00
Chemists	SIOP Fellows	47.93	9.61	.01	32.55	8.10	0.02
Chemists	SIOP Members	58.53	9.51	.00	46.40	7.61	0.00
SIOP Fellows	Psychology				-86.79	20.71	0.02
SIOP Fellows	Economics				-68.45	15.45	0.02
SIOP Fellows	Physicists	-30.63	4.54	.00	-56.56	9.05	0.00
SIOP Fellows	SIOP Members	10.60	1.65	.00	13.86	3.12	0.01
SIOP Fellows	Chemists	-47.93	9.61	.01	-32.55	8.10	0.02

Note: only significant differences at $p < .05$ displayed.

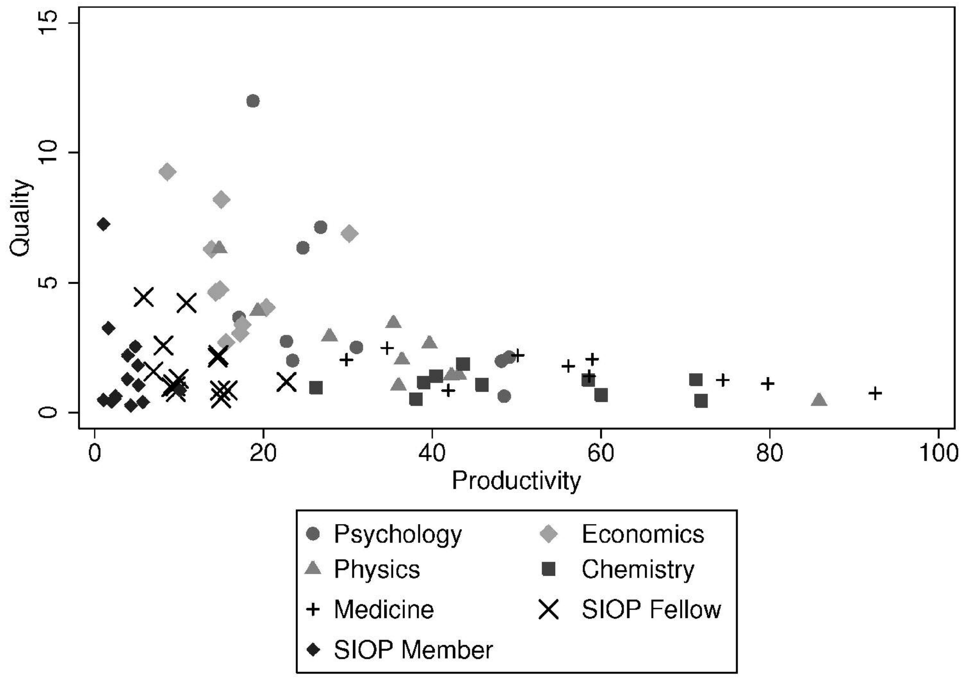

Figure 1: Relationship between Quality and Productivity
139x101mm (600 x 600 DPI)