



Does a Long Reference List Guarantee More Citations? Analysis of Malaysian Highly Cited and Review Papers

Nader Ale Ebrahim¹, H. Ebrahimian², Maryam Mousavi³, Farzad Tahiri³

¹Research Support Unit, Centre of Research Services, Institute of Research Management and Monitoring (IPPP), University of Malaya, Malaysia

²Institute of Mathematical Sciences, Faculty Science, University Malaya.

³Centre for Product Design and Manufacturing, Department of Mechanical Engineering, Faculty of Engineering, University of Malaya, 50603, Kuala Lumpur, Malaysia

Corresponding author (e-mail): alebrahim@um.edu.my

Abstract: Earlier publications have shown that the number of references as well as the number of received citations are field-dependent. Consequently, a long reference list may lead to more citations. The purpose of this article is to study the concrete relationship between number of references and citation counts. This article tries to find an answer for the concrete case of Malaysian highly cited papers and Malaysian review papers. Malaysian paper is a paper with at least one Malaysian affiliation. A total of 2466 papers consisting of two sets, namely 1966 review papers and 500 highly-cited articles, are studied. The statistical analysis shows that an increase in the number of references leads to a slight increase in the number of citations. Yet, this increase is not statistically significant. Therefore, a researcher should not try to increase the number of received citations by artificially increasing the number of references.

Key words: H-index, Citation analysis, Bibliometrics, Impact factor, Performance evaluation, Relations between citations and references

1. Introduction

Researchers seeking citation tracking to find the most influential articles for a particular topic and to see how often their own published papers are cited (Bakkalbasi et al. 2006). On the other hand universities are looking for citations because of its influence in the university ranking (Ale Ebrahim et al. 2013, Ioannidis 2010, Bornmann, Leydesdorff, and Wang 2014). A citation count is the number of times a research work such as a journal article is cited by other works. The citation per paper meaningfully influence a number of metrics, including total citation counts, citation speed, the ratio of external to internal cites, diffusion scores and h-index (Carley, Porter, and Youtie 2013). Citation counts still commonly use for the measure of research papers quality and reputation (Abt and Garfield 2002). The number of citations that an article receives measured its impact on a specific field (Lai, Darius, and Lerut 2012). Citation analysis is one of the most important tools to evaluate research performance (Bornmann et al. 2012). Citation indicator is important for scientists and universities in all over the world (Farhadi, Salehi, Yunus, et al. 2013). In the early stage, the relationship between the number of references and the number of the paper citation was investigated in the 1965 (UZUN 2006, de Solla Price 1965). A long reference list at the end of a research paper may be the key to ensuring that it is well cited (Corbyn 2010, Ball 2008). Hence, citation counts are correlated with reference frequencies (Abt and Garfield 2002). Webster, Jonason, and Schember (2009) raised the question "Does the number of references an article contains predict its citation count?" and found that reference counts explained 19% of the variance in the citation counts. Lancho-Barrantes, Guerrero-Bote, and Moya-Anegón (2010) found that not only the number, but also the citation impact of the cited references correlated with the citation counts for a paper. The higher the impact of the cited references, the higher the later impact of the citing paper (Bornmann et al. 2012). Review articles are usually highly cited compare to other types of papers (Meho 2007).

Review papers represent the existing knowledge in a given field and more likely to be cited (Alimohammadi and Sajjadi 2009). Several bibliometric studies highlighted that citation counts are a function of many factors besides the scientific quality (Bornmann et al. 2012), length of paper (Abt and Garfield 2002), visibility (Ale Ebrahim et al. 2014), optimize scholarly literature for academic search engines (Beel, Gipp, and Wilde 2010), add the name of study in the title of all publications

(Sarli and Holmes 2011), publishing in a journal with higher impact factor (Vanclay 2013), internet usage (Farhadi, Salehi, Embi, et al. 2013), gross domestic product (GDP) (Gholizadeh et al. 2014), number of authors (Krause 2009), self-archiving (Gargouri et al. 2010), publish in an open access journal (Swan 2010), collaborate with international authors (Pislyakov and Shukshina 2012), write paper with a Nobel laureates (Ball 2011) and many other (Ale Ebrahim et al. 2013) including write a review paper (Vanclay 2013) and use more references (Corbyn 2010). In this study the relationship between number of references and citation counts is determined. Webster, Jonason, and Schember (2009) mentioned "On average, review articles actually showed less of the relationship than standard articles" (Corbyn 2010). So, in this research both review and standard articles (papers) were investigated. 2466 articles consist of 1966 Malaysian review and 500 highly cited papers were selected to examine the relationship between number of references and citation counts in the given article.

2. Materials and methods

All data were obtained through Web of Science online academic database provided by Thomson Scientific. This database included the necessary information to examine the relationship between reference and citation counts for every review and highly cited papers published in Malaysia since 1980 to October 2013. Science Citation Index Expanded, Social Sciences Citation Index and Arts & Humanities Citation Index, were searched for reviews and highly cited papers. For each paper, all Bibliometrics data, especially the number of references and the number of times the paper has been cited during the interval between the year of publication and the year 2013, have been collected. Two samples set were selected: 1- The sample number one consisted of 1966 review papers in all disciplines from Malaysia, according to the Web of Knowledge's classification system. Citation statistics produced by shorter than three years' time frame may not be sufficiently stable (Adams 2005, UZUN 2006). Because, papers appearing in the Web of Science databases over the last few years, have not had enough time to accumulate a stable number of citations (Webster, Jonason, and Schember 2009). Therefore, the time span limited from 1980 to November, 2010; yielding a subsample of 721 publications (37% of the original sample). Publications with zero citation were removed. In order to select the highly cited paper a threshold 10 times cited per year is

considered. The association between the number of references (independent variable) and time cited per year (dependent variable) of highly cited review papers investigated with linear and non-linear models. 2- The sample number two comprises 500 highly cited publications from Malaysia. According to the Web Of Science classification, the results are obtained based on the article type and exclude the review articles, editorial material, conference papers and book review.

3. Results and discussion

Two sets of data 1- 1966 review papers and 2- 500 high cited papers, were investigated separately. The results and discussions are coming as follows.

Outliers for sample one (1966 review papers)

Due to the effect of the age of an article, the number of citations cannot be a reference of highly cited paper. Therefore, the citation per year selected as a reference for highly cited paper. Papers with 10 times cited per year is considered as highly cited paper. Figure 3-1 shows the number of times cited per year for 660 review papers. A threshold was visually determined on 50 times cited per year. Papers with more than 50 times cited yearly is called “extremely high cited paper” and detected as outliers. Papers with more than 300 listed references also detected as outliers (3-2).

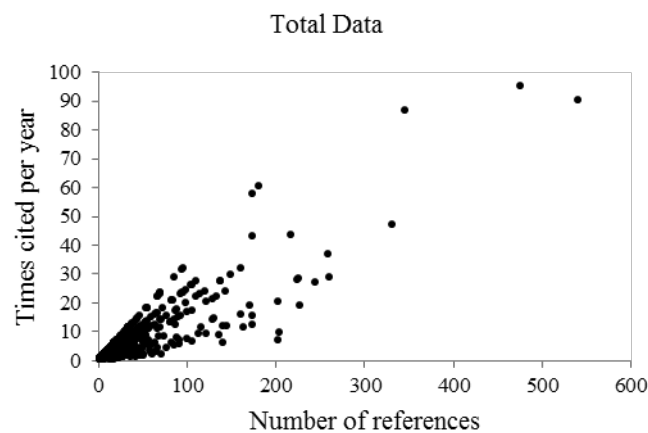


Figure 3-1 Number of times cited per year vs number of review papers references

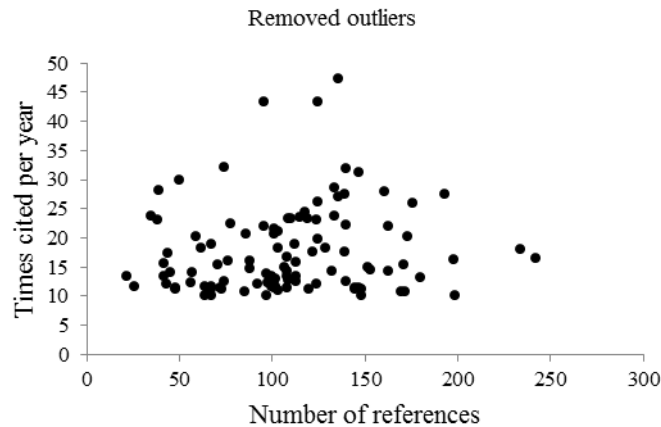


Figure 3-2 Number of times cited per year vs number of references in review paper

Correlation analysis for sample one (1966 review papers)

The correlation between variables was modeled with regression model, linear model $y = \alpha x + \beta$ and exponential model, non-linear model $y = \alpha e^{\beta x}$. The goodness of both model was then measured with Spearman's rho (ρ), Kendall's tau (τ) and Pearson correlation coefficient (r). The result of correlation analysis is summarized in 3-1.

Coefficient	Statistic value	P-value	Equation	Model
ρ	0.127	0.193	$y = 0.0639x + 10$	Linear
τ	0.112	0.093	$y = 10e^{0.0041x}$	
r	0.152	0.120		Non-linear

Table 3-1 The result of correlation analysis of highly-cited review papers

The association between variables is graphically illustrated with scatter plots. The trend of these associations was drawn with solid lines. Refer to Figure 3 and Figure 4, both linear and non-linear models are not significantly fitted, trends are positive which support the hypothesis "For a given review paper, increasing in the number of references may have result of increasing the times cited per year".

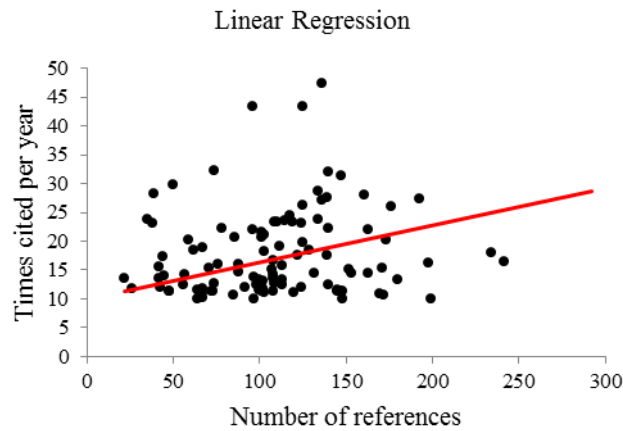


Figure 2-3 Relationship between number of references and citation counts in review papers (linear model)

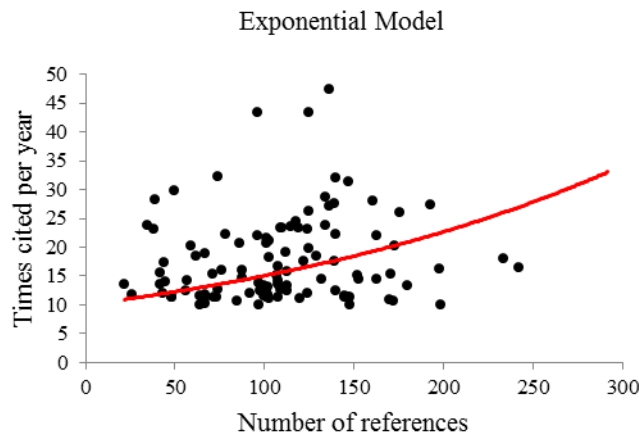


Figure 3-3 Relationship between number of references and citation counts in review papers (Exponential model)

Outlier detection for sample two (500 highly cited papers)

Papers with 10 times cited per year is considered as highly cited paper. Papers that cited more than 100 times per year is considered as extremely high cited paper and detected as an outlier. Figure 5 and Figure 6 are showing raw data and filtered data respectively.

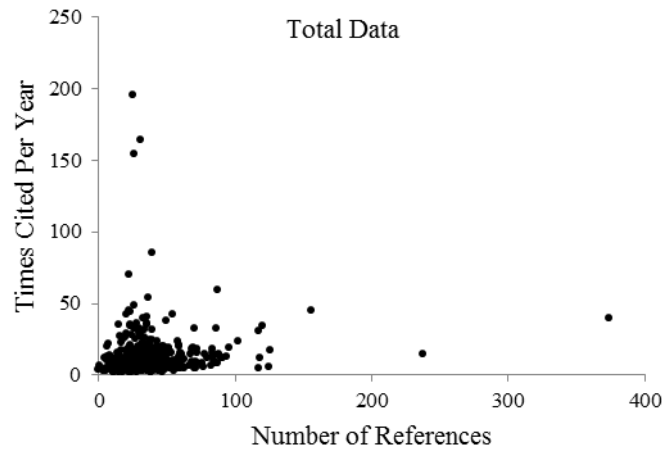


Figure 3-4 Raw data - Number of times cited per year vs number of references 500 highly cited papers

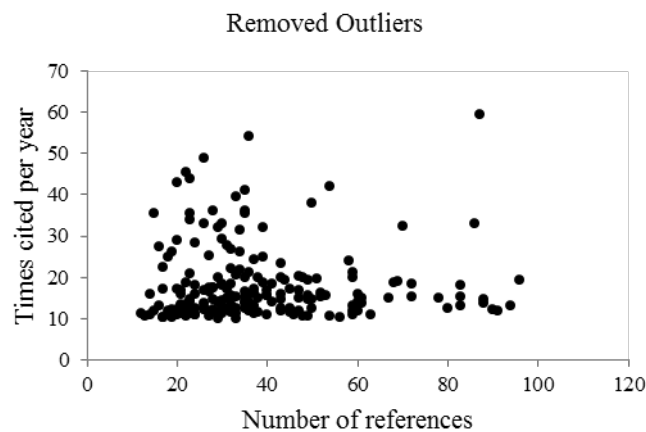


Figure 3-5 Filtered data - Number of times cited per year vs number of references in 500 highly cited papers

Correlation analysis for sample two (500 highly cited papers)

The association between the number of references (independent variable) and time cited per year (dependent variable) of first 500 high cited papers investigated with linear and non-linear model correlation analysis. The correlation was modeled with regression model, linear model $y = \alpha x + \beta$ and exponential model, non-linear model $y = \alpha e^{\beta x}$. The goodness of fit was then measured with Spearman's rho (ρ), Kendall's tau (τ) and Pearson correlation coefficient (r). The result of correlation analysis is summarized in Table 3-2.

Coefficient	Statistic value	P-value	Equation	Model
ρ	0.08	0.914	$y = 0.03x + 2.4$	Linear
τ	0.04	0.467	$y = 1.1 e^{0.08x}$	
r	0.06	0.419		Non-linear

Table 3-2 The result of correlation analysis of 500 highly cited papers

The association between variables is graphically illustrated with scatter plots. The trend of these associations is shown by the solid lines. Figure 3-7 and Figure 3-8 shows, although both linear and non-linear models are not significantly fitted, positive values of correlation coefficients are still suggesting a positive trend (positive correlation) on the number of references and the number of times cited per year.

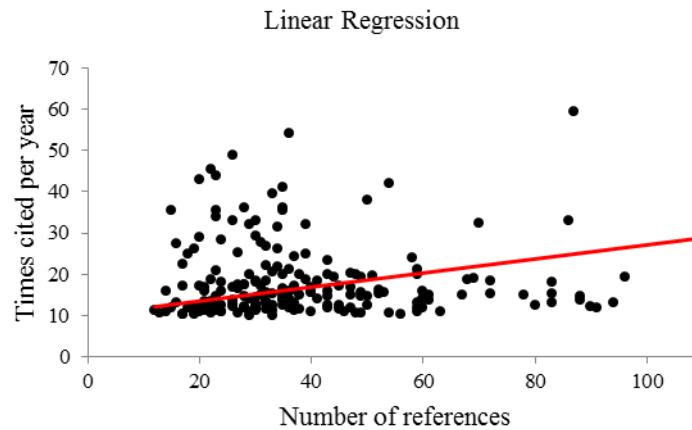


Figure 3-6 Relationship between number of references and citation counts in 500 highly cited (linear model)

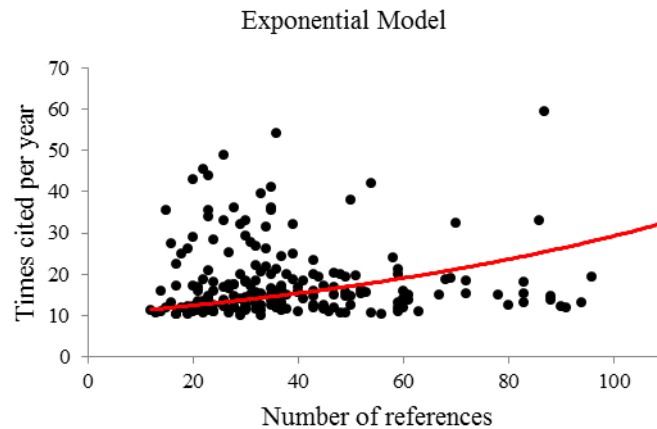


Figure 3-7 Relationship between number of references and citation counts in 500 highly cited (Exponential Model)

4. Conclusion

This study shows that since the trend between the citation count and the number of references is not statistically significant, we cannot conclude that there is a significant association between the citation count of Malaysia review papers between the given period and number of references contained in the paper. The correlation coefficient is not statistically significant. However, $r = 0.152$ based on the population of 721 articles. Malaysian review papers get more citations than other types of papers. The number of references in the article has the lowest impact on the citation compares with review paper. As this study looked only Malaysia review papers and 500 highly-cited article, it would be necessary to conduct a similar study in the otherworld and types of papers. It would be important to examine whether in other types of papers the relationship investigated here have significant correlated or not. The research considered the general definition of citations. Therefore, future studies may make a differentiation between “perfunctory citations” and “organic citations” citations as Tang and Safer (2008) defined “perfunctory citations” is occurred only once and in the introduction, “organic citations” as references cited for “conceptual ideas” and “methodology and data” reasons.

ACKNOWLEDGEMENT

Sincere thanks to Dr. Bojan Obrenović and the International Journal of Management Science and Business Administration's board members for their useful advices.

References

- Abt, Helmut A., and Eugene Garfield. 2002. "Is the relationship between numbers of references and paper lengths the same for all sciences?" *Journal of the American Society for Information Science and Technology* 53 (13):1106-1112. doi: 10.1002/asi.10151.
- Adams, Jonathan. 2005. "Early citation counts correlate with accumulated impact." *Scientometrics* 63 (3):567-581. doi: 10.1007/s11192-005-0228-9.
- Ale Ebrahim, Nader, Hadi Salehi, Mohamed Amin Embi, Farid Habibi Tanha, Hossein Gholizadeh, and Seyed Mohammad Motahar. 2014. "Visibility and Citation Impact." *International Education Studies* 7 (4):120-125. doi: 10.5539/ies.v7n4p120.
- Ale Ebrahim, Nader, Hadi Salehi, Mohamed Amin Embi, Farid Habibi Tanha, Hossein Gholizadeh, Seyed Mohammad Motahar, and Ali Ordi. 2013. "Effective Strategies for Increasing Citation Frequency." *International Education Studies* 6 (11):93-99. doi: 10.5539/ies.v6n11p93.
- Alimohammadi, Dariush, and Mahshid Sajjadi. 2009. "Correlation between references and citations." *Webology* 6 (2):a71.
- Bakkalbasi, Nisa, Kathleen Bauer, Janis Glover, and Lei Wang. 2006. "Three options for citation tracking: Google Scholar, Scopus and Web of Science." *Biomedical Digital Libraries* 3 (1):7. doi: 10.1186/1742-5581-3-7.
- Ball, Philip. 2008. "A longer paper gathers more citations." *Nature* 455 (7211):274-275. doi: 10.1038/455274a.
- Ball, Philip. 2011. "Are scientific reputations boosted artificially?" *Nature*, 6 May.
- Beel, Jöran, Bela Gipp, and Erik Wilde. 2010. "Academic Search Engine Optimization (ASEO)." *Journal of Scholarly Publishing* 41 (2):176-190. doi: 10.3138/jsp.41.2.176.
- Bornmann, L., L. Leydesdorff, and J. Wang. 2014. "How to improve the prediction based on citation impact percentiles for years shortly after the publication date?" *Journal of Informetrics* 8 (1):175-180. doi: 10.1016/j.joi.2013.11.005.
- Bornmann, Lutz, Hermann Schier, Werner Marx, and Hans-Dieter Daniel. 2012. "What factors determine citation counts of publications in chemistry besides their quality?" *Journal of Informetrics* 6 (1):11-18. doi: <http://dx.doi.org/10.1016/j.joi.2011.08.004>.
- Carley, S., A. L. Porter, and J. Youtie. 2013. "Toward a more precise definition of self-citation." *Scientometrics* 94 (2):777-780. doi: 10.1007/s11192-012-0745-2.
- Corbyn, Zoë. 2010. "An easy way to boost a paper's citations." *Nature* 406. doi: 10.1038/news.2010.406
- de Solla Price, Derek J. 1965. "Networks of Scientific Papers." *Science* 149 (3683):510-515. doi: 10.1126/science.149.3683.510.
- Farhadi, Hadi, Hadi Salehi, Melor Md Yunus, Arezoo Aghaei Chadegani, Maryam Farhadi, Masood Fooladi, and Nader Ale Ebrahim. 2013. "Does it Matter Which Citation Tool is Used to Compare the h-index of a Group of Highly Cited Researchers?" *Australian Journal of Basic and*

- Applied Sciences 7 (4):198-202. doi: arXiv:1306.0727.
- Farhadi, Maryam, Hadi Salehi, Mohamed Amin Embi, Masood Fooladi, Hadi Farhadi, Arezoo Aghaei Chadegani, and Nader Ale Ebrahim. 2013. "Contribution of Information and Communication Technology (ICT) in Country'S H-Index." *Journal of Theoretical and Applied Information Technology* 57 (1):122-127. doi: 10.5281/zenodo.7715.
 - Gargouri, Yassine, Chawki Hajjem, Vincent Larivière, Yves Gingras, Les Carr, Tim Brody, and Stevan Harnad. 2010. "Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research." *PLoS ONE* 5 (10):e13636. doi: 10.1371/journal.pone.0013636.
 - Gholizadeh, Hossein, Hadi Salehi, Mohamed Amin Embi, Mahmoud Danaee, Seyed Mohammad Motahar, Nader Ale Ebrahim, Farid Habibi Tanha, and Noor Azuan Abu Osman. 2014. "Relationship among Economic Growth, Internet Usage and Publication Productivity: Comparison among ASEAN and World's Best Countries." *Modern Applied Science* 8 (2):160-170. doi: 10.5539/mas.v8n2p160.
 - Ioannidis, J. P. A. 2010. "Is there a glass ceiling for highly cited scientists at the top of research universities?" *Faseb Journal* 24 (12):4635-4638. doi: 10.1096/fj.10-162974.
 - Krause, Kate. 2009. Increasing your Article's Citation Rates. Open Access Week. Accessed 28 May 2013.
 - Lai, Quirino, Tom Darius, and Jan Lerut. 2012. "The one hundred most frequently cited articles in the field of clinical liver transplantation." *Transplant International* 25 (6):e76-e77.
 - Lancho-Barrantes, Bárbara S., Vicente P. Guerrero-Bote, and Félix Moya-Anegón. 2010. "What lies behind the averages and significance of citation indicators in different disciplines?" *Journal of Information Science* 36 (3):371-382. doi: 10.1177/0165551510366077.
 - Meho, Lokman I. 2007. "The rise and rise of citation analysis." *Physics World* 20:32-36.
 - Pisyakov, Vladimir, and Elena Shukshina. 2012. "Measuring Excellence in Russia: Highly Cited Papers, Leading Institutions, Patterns of National and International Collaboration." *Proceedings of STI 2012, Montréal*.
 - Sarli, Cathy, and Kristi Holmes. 2011. "Strategies for Enhancing the Impact of Research." Washington University School of Medicine, Accessed 9 May. <https://becker.wustl.edu/impact-assessment/strategies>.
 - Swan, Alma. 2010.
 - Tang, Rong, and Martin A. Safer. 2008. "Author-rated importance of cited references in biology and psychology publications." *Journal of Documentation* 64 (2):246-272. doi: 10.1108/00220410810858047.
 - Uzun, Ali. 2006. "Statistical relationship of some basic bibliometric indicators in scientometrics research." *International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting, Nancy (France), May 10 - 12, 2006*.
 - Vanclay, Jerome K. 2013. "Factors affecting citation rates in environmental science." *Journal of Informetrics* 7 (2):265-271. doi: <http://dx.doi.org/10.1016/j.joi.2012.11.009>.
 - Webster, Gregory D, Peter K Jonason, and Tatiana Orozco Schember. 2009. "Hot topics and popular papers in evolutionary psychology: Analyses of title words and citation counts in *Evolution and Human Behavior*, 1979-2008." *Evolutionary Psychology* 7 (3):348-362.