Assessment Word Count ______11697_________. Coursework submitted after the maximum period will receive zero marks. Your assignment has a word count limit. A deduction of 3 marks will be applied for coursework that is 5% or more above or below the word count as specified above or that does not state the word count.

Ethics documentation is included in the Appendix if your dissertation has been judged to be Low Risk or High Risk. ✓ (Please tick the box if you have included the documentation) A deduction of 3 marks will be applied for a dissertation if the required ethics documentation is not included in the appendix. The deduction procedures are detailed in the INF6000 Module Outline and Dissertation Handbook (for postgraduates) or the INF315 Module Outline and Dissertation Handbook (for undergraduates).
Access to Dissertation

A Dissertation submitted to the University may be held by the Department (or School) within which the Dissertation was undertaken and made available for borrowing or consultation in accordance with University Regulations.

Requests for the loan of dissertations may be received from libraries in the UK and overseas. The Department may also receive requests from other organisations, as well as individuals. The conservation of the original dissertation is better assured if the Department and/or Library can fulfill such requests by sending a copy. The Department may also make your dissertation available via its web pages.

In certain cases where confidentiality of information is concerned, if either the author or the supervisor so requests, the Department will withhold the dissertation from loan or consultation for the period specified below. Where no such restriction is in force, the Department may also deposit the Dissertation in the University of Sheffield Library.

To be completed by the Author – Select (a) or (b) by placing a tick in the appropriate box

If you are willing to give permission for the Information School to make your dissertation available in these ways, please complete the following:

- (a) Subject to the General Regulation on Intellectual Property, I, the author, agree to this dissertation being made immediately available through the Department and/or University Library for consultation, and for the Department and/or Library to reproduce this dissertation in whole or part in order to supply single copies for the purpose of research or private study

- (b) Subject to the General Regulation on Intellectual Property, I, the author, request that this dissertation be withheld from loan, consultation or reproduction for a period of [ ] years from the date of its submission. Subsequent to this period, I agree to this dissertation being made available through the Department and/or University Library for consultation, and for the Department and/or Library to reproduce this dissertation in whole or part in order to supply single copies for the purpose of research or private study

Name: Zhen Fan
Department: Information Systems
Signed: Zhen Fan
Date: 01/09/2015

To be completed by the Supervisor – Select (a) or (b) by placing a tick in the appropriate box

- (a) I, the supervisor, agree to this dissertation being made immediately available through the Department and/or University Library for loan or consultation, subject to any special restrictions (*) agreed with external organisations as part of a collaborative project.

- (b) I, the supervisor, request that this dissertation be withheld from loan, consultation or reproduction for a period of [ ] years from the date of its submission. Subsequent to this period, I, agree to this dissertation being made available through the Department and/or University Library for loan or consultation, subject to any special restrictions (*) agreed with external organisations as part of a collaborative project.

Name
Department
Signed
Date

THIS SHEET MUST BE SUBMITTED WITH DISSERTATIONS BY DEPARTMENTAL REQUIREMENTS.
A study of the extent to which citation counts correlate with authors’ perception of the importance of a publication

A study submitted in partial fulfillment of the requirements for the degree of Msc Information Systems

at

THE UNIVERSITY OF SHEFFIELD

by

Zhen Fan

September 2015
Abstract

**Background.** The literature reveals that although citation counts was common applied on evaluation of authors, institution, countries as well as articles, the extent to which citation counts could reflect the real quality of an article is seldom explored. Previous studies indicated many factors that influence the citation counts, which revealed the limitations of using citation counts in academic evaluation.

**Aims.** This study aims to investigate the extent to which authors’ perceptions of the importance of their publications are correlated with citation counts of them. In addition, the reasons and the phenomenon behind the correlation as well as the processes how some factors making influence on the results are tried to be explored in this study.

**Methods.** The correlation was indicated by comparing regarded important publications with highly cited publications of an author. The important publications were collection through emails with authors. Then the reasons and phenomenon were explored by face to face interviews with authors. 8 of academic staff from information school of University of Sheffield consented to participate this study and provided total 76 publications as their important publications.

**Results.** With analysis of measures in three stages, the results of correlation in this study are 68.42%, 71% and 69.13%. These results indicate that the correlation between authors’ perception of important and citation counts is moderate. In further exploration of this study, publishing year, publishing place and type of publications are the external factors that have significant influence on the citation counts. Furthermore, the analysis of reasons why these publications were regarded as important shows that career life and personal experience are the main reasons which not relate to the quality of publications.

**Conclusions.** The moderate correlation in this study shows that citation counts could reflect the quality of a publication to some degree with there are some external factors need to be considered. The results of the reasons of important suggested that the concept of important is probably has more personal meaning. Which might be the main reasons that cause the difference between citation counts and author’s perception of importance of a publication.
Acknowledgement

This dissertation would not been completed without the help of many individuals. I would like to give my appreciation to the following people.

Firstly, I would like to give my thanks to my supervisor, Peter Willett. His support and advice were very meaningful to this study.

Then I would like also to appreciate the academic staff who have participated my study. Without their generous help, I would not collect data and complete this dissertation.
# Contents

Abstract ................................................................................................................................. 5  
Acknowledgement ................................................................................................................ 7  
1 Introduction ......................................................................................................................... 9  
   1.1 Background .................................................................................................................... 9  
   1.2 Aims and objectives ....................................................................................................... 9  
   1.3 Structure of dissertation ............................................................................................... 9  
2 Literature review .................................................................................................................. 9  
   2.1 Studies on using citation counts to evaluate the importance of publication and relevant issues 9  
   2.2 Studies on subjective factors which influence the citation counts of an article ....... 10  
   2.3 Studies on subjective factors which influence the citation counts of an article ....... 11  
   2.4 Studies on features concerning article content as factors ........................................ 11  
   2.5 Studies using self-evaluation ....................................................................................... 11  
3 Methodology ....................................................................................................................... 12  
   3.1 Research approaches ..................................................................................................... 12  
   3.2 Methods of Investigation ............................................................................................. 12  
      3.2.2 Interviews Preparation ......................................................................................... 12  
      3.2.3 Interview guide design ....................................................................................... 13  
      3.2.4 Data Recording .................................................................................................... 13  
   3.3 Results analysis ............................................................................................................. 13  
   3.4 Ethics issues .................................................................................................................. 13  
4 Findings ............................................................................................................................... 13  
5 Discussion ........................................................................................................................... 32  
6 Conclusion .......................................................................................................................... 34  
Reference .................................................................................................................................. 36  
Appendix 1 ............................................................................................................................. 38  
Appendix 2: Consent Form .................................................................................................... 43  
Appendix 3: Approval Letter ................................................................................................ 45
1 Introduction

1.1 Background

Using citation counts to evaluation the importance of a publication is a common application which has attracted increasing attention these years (Moed, 2005). Citation counts was initially used as a quantitative measure to evaluate the importance of an article. Now it is common used on evaluation of authors, institution, countries as well as articles. There is no doubt that citation counts could reveal the impact of an article in statistics effectively. However, this measure also faces with many criticisms. Citation counts could only reflect the utility and impact of an article, which is a valuable part of indicators to evaluate an article. Many previous studies indicated that the correlation between citation counts and peer evaluation is only approximately 0.4-0.6 (Lutz and Daniel, 2006), which means that an article’s citation counts is not always agree with its common evaluation. The limitations must be considered when using citation counts in academic evaluation.

Reasons lead to this situation might be that, besides the quality of a work, there are some objective features of authors, articles and journal influence the citation, which has been demonstrated by a good deal of studies. Moreover, some subjective factors of citers including their unconventional motivation of citation (citing as negative example, citing to bedding their own researches, etc.) and low quality of citation also need to be taken into consideration. Since many objective factors (external factors) have been proved as significant influence factors, effect of subjective factors (internal factors) are usually low. Therefore, it is no need to be regarded as a serious problem.

As there are many studies gain efficiency on how external factors mentioned above affect the validity of this evaluation approach, very few studies have examined how this evaluation results are consistent to the real importance of publication. This circumstance might be resulted by the difficulty of non-experts to understanding the significance of scientific papers (Small, 2004). One solution could be to compare the authors’ self-assessment on importance of publication with the real citation counts of their articles, for authors themselves would be the person who is most familiar with their works as well as their study fields.

1.2 Aims and objectives

The aim of this study is to examine whether the citation counts for a research article correlates to authors’ self-evaluation on the importance of the articles. This study investigated self-assessment of authors through face-to-face interviews with academic staff from information school (iSchool) in University of Sheffield as authors of many works in similar fields. The similarity of research fields, research environment as well as the same institution they affiliated would reduce the influence of external factors on exploring of questions.

At the beginning of the data collection, invitations were sent to staff of the iSchool to participate in the study. They were asked to provide a list of their most important articles (up to ten). Then, citation counts for those articles were identified in the Web of Science and Google Scholar. The interviews were designed to discuss why the chosen publications are important, why some unimportant publications were highly cited and the possible reasons of resulting correlation. The results of the citation counts were used to analysis the extent of the correlation. Then the records of interviews are used to discuss and explore the reasons of the situation of which the extent of consistency is not remarkable or which shows negative correlation.

1.3 Structure of dissertation

This dissertation will present results of literature review of previous studies on using citation counts to evaluation the importance and its relevant issues, the studies on factors (both external and internal) which may affect the citation counts and the studies using peer evaluation and self-evaluation to measure the importance of articles. The methodology of this study are defined according to the knowledge of these previous studies.

Then the results of this study were presented basing on data collecting. The questions of interviews are summarized in 5 questions below:

1. The extent to which citation counts correlate with authors’ perception of importance of a publication.
2. Reasons of publications are regarded as important.
3. Authors’ perception about why their important publications are not highly cited.
4. Reasons of highly cited publications were not regarded as important and reasons why they are cited.
5. Inconsistency between importance and citation counts among the important publications.

The content of discussion chapter is basing on the results of findings chapter. In discussion chapter, this study try to explore the reasons and the phenomenon behind the results. Additionally, try to find how some factors made influence on the citation counts.

In conclusion chapter, all results of this study are summarized. Suggestions for further studies and limitations of this study are also included in this chapter.

2 Literature review

2.1 Studies on using citation counts to evaluate the importance of publication and relevant issues
Early in 1927, in paper published by Gross and Gross (1927), citation counts is used to evaluate the importance of scientific work (Lutz and Daniel, 2008). Which pointed out a method to use the number of references for a college library deciding what files of scientific periodicals to purchase without experts involved. Since then, there are trends to develop evaluation systems to identify excellent universities, research groups and researchers in many countries (Danell 2011, p. 50). On the other hand, citation also used to rank lists of journals, papers and educational institutions (Bar-Ilan, 2008). In these studies the citation counts were used to measure the impact of the work, as high quality work will obtain more citation from scientific colleagues than low quality work (Van Raan, 2003). This methods of evaluation also faced with many criticisms. A series of researches of MacRoberts and MacRoberts (1987, 1989, 1996, 2010) pointed out that the basic assumption of evaluation that the research cited represents a roughly valid indication of influence no their work has been falsified repeatedly. ‘Citing is biased, secondary sources replace primary sources, and informal sources, which are the lion’s share, are not credited (1996).’ Their study of uncited and seldom-cited influences indicated the results that uncited work should be taken into account in evaluative citation analysis (2010). Lindsey (1989) also examined four kinds of limitations which proved to be problems involved in the use of citation counts as a measure of quality in science.

It is also not clear how the range of excellent papers should be defined (Glanzel and Schubert 1992; Kostoff, Barth & Lau, 2008). There are basically two different approaches to define what counts as a highly cited papers, involving absolute or relative thresholds (Aksens, 2003, pp.160). In Lutz’s study (2008) on how to define an excellent paper, absolute number was use in most of the papers. Definitions based on a relative number should however be preferred to definitions based on an absolute number, as only they can be used for cross-field and cross-time-period comparisons. This study also indicate that around half papers use percentile rank classed to identify the excellent papers, which appears to be the preferred method with which to identify excellent papers (Bornmann, 2013). The percentile rank classes used to identify excellent papers are different. Adams, Pendlebury and Stembridge (2013) use the world’s top 1% most cited papers in that year to study the share of them. The Leiden Group has introduced the percentage in the top 5% of cited papers of as an indicator to evaluate the excellence (Albarran, Ortufno & Ruiz-Castillo, 2011). Even so, citation counts still continues to be widely and increasingly used as measures of evaluation.

2.2 Studies on subjective factors which influence the citation counts of an article

Various subjective factors have been demonstrated by many studies that may influence the citation counts of an article. Some recent studies and their results are listed below.

Natsuo and Fuyuki (2014) have analyzed original articles published in the same year in six selected subject field. The results shows some generality across the fields is that the Price Index (percentage of the references whose publication year is within 5 years before the publication year of the article) was the strongest predictor of citation and number of references was the next. The effects of number of authors and authors’ achievement measures were rather weak.

Didegah and Thelwall (2013) using over 50,000 articles published from 2007 to 2009 in the field of nano-science and nanotechnology. Among 8 factors they selected, the journal impact factor (JIF) and impact of references (the mean citations of referenced publications) show high significant as an influence factors while number of countries of affiliation shows have a negative effect on citation counts. Moreover, the number of references, internationality of references and number of institution of affiliation were also influence the citations in a significant level.

Peng and Zhu’s (2012) study among 18,580 social science articles on Internet studies also indicated a strong effect of journal characteristics (for example, JIF). Other significant predictors including some article characteristics such as article length, number of authors, topical popularity, the pro-portio of highly cited publications in references and active years of the first author (author characteristics).

Haslam et al. (2008) analysed the citation counts of 308 articles published in three major journals of social-personality psychology in 1996. Thirty potential factors were used as the explanatory variables and were classified into four groups (characteristics of author, institution, article organization and research approach). Then the nine significant variables in first stage were used as explanatory variables in next stage. This research indicated that the significant positive factors were: number of past publications of the first author; high productivity of a coauthor; high journal prestige; number of pages in the article; number of references and recency of references.

Walters (2006) used 9 explanatory variables to predict the citation counts of 428 articles from 12 prime psychology journals in 2003. The results shows that the positive influence of the first author’s past publications, the first author’s nationality and whether it was a review article is significant. According to the results, author factors might be more important than journal and article factors.

Van Dalen and Henkens (2001, 2005) focused on the roles of author and journal reputation and counted citations received by 1,371 articles in the field of demography with the citation windows of 5 years and 10 years after publication. Among the variables, both journal reputation and author reputation shows significant influence. Other significant factors were article type, number of pages, regional focus of the article, and the language of the journal.

Fu and Aliferis (2010) develop a learning model which was applied to 3,788 article sampled from 8 general medical
journals published between 1991 and 1994. The effective variables were selected from content-based features and bibliometric features. Among bibliometric features, the JIF accumulated number of citations obtained by the last author were significant. On the other hand, when the value of the citation count changed, the effective content-based features varied greatly. According to the studies below and many other studies are not listed, external factors have significant influence on citation counts of an article. Which could indicate from one side that the veracity of evaluation the importance of an article using citation counts is questionable.

2.3 Studies on subjective factors which influence the citation counts of an article
Besides the influence of objective data, citers’ motivation and citation quality could also act as influencing factors of citations. Authors’ motivations were firstly analyzed in Brooks’ study (1985). According to which persuasiveness appeared to be the major motivator and the influence of social consensus or negative credit was very weak. Brooks’ further study (1986) found that the majority of the references were attributed to more than one motive. Vinkler (1987) proposed a quasi-quantitative citation model for citation behavior and found that professional motivations was the predominant reason for citing while connectional reasons is very weak. There are also problems on identifying the motivation of citing. Willett (2012) found in his studied that reader-reasons and author-reasons for citing are typically so different while authors’ reasons are sometimes difficult to obtain accurately. Hence the authors’ motivations are not easy to be recognized.

In aspect of citation quality, Liu’s study (1993) on citing practice of Chinese physicists shows that only a minority of scientists said that more than 80% of their citations were essential. Moravcsik and Muregesan (1975) analyzed the quality of citation by classifying citation into four dimensions: conceptual or operational; organic or perfunctory; evolutionary or juxtapositional; confirmative or negational. A large fraction of references were found perfunctory, which raised serious doubts about the use of citations as a quality measure. According to studies above, the effect of citers’ motivation is weak but is not easy to identify which the citation quality might be a serious problems on using citation counts as a measure of evaluation. Moreover, these internal factors are not easy to be recognized intuitively in studies.

2.4 Studies on features concerning article content as factors
Some of the studies involved features concerning article content as factors. Natsuo and Fuyuki (2014) summarized these features include topic terms of medical articles (Fu & Aliferis, 2010), research design of clinical medicine research (Lokker et al., 2008), subfields and research methods of geoscience research (Stewart, 1983), and themes of demo-graphic articles (van Dalen & Henkens, 2001, 2005).

2.5 Studies using self-evaluation
There are some previous studies which use self-evaluation by the authors on their publications. Measures and results of these studies provided valuable experience in developing the idea and methodology of current study. In 1987, Porter et al. compared various bibliometric measures and scientists’ own judgments. This study found that the extent of overlap between Best and Most-cited papers is about 1/3, which is too small to equating the highly cited papers to papers were judged best by authors. In this study, three “best” papers were nominated by Sloan Fellows in Chemistry. For those “best” and those accruing most citations, detailed information were compiled on the exact year and publication of each citation. In addition, interviews were used to probe authors’ perceptions of scientific progress and third-person peer evaluations were used to cast light on the meritorious scientific contribution. Other data from behalf of nominees for Lawrence Awards and other scientists were also used to presented analyses. Besides the extent of overlap between best and most-cited papers, this study also explored the difference by type of paper, the citation patterns and cross-disciplinary exchange. Small (2004) used self-evaluation to investigate the reasons why authors think their papers are highly cited. These self-perceptions provided clues to the factors that lead to high citation rate and the importance of the interaction between internal and external factors. Data was collected through questionnaire to the authors of papers in each of the 22 fields on a monthly basis. The qualitative analysis of the returns suggests four main dimensions: interest, novelty, utility and significance. From the frequency analysis of reasons and co-occurrence analysis, interest is the most common component to the explanation of high citation rate. In Aksnes’s study (2006) about citation rates and perceptions of scientific contribution, scientists were asked about their own publication history and their citation counts. The results showed that the citation
counts correspond well with the authors’ perceptions of scientific contribution. In his study, 297 highly cited articles published by Norwegian scientists in the 15-year period (1981-1996) were included. Questionnaires were sent to the identified 221 authors of these articles and obtained 166 responded (75% response rate). In the questionnaires, authors were asked to consider the paper’s scientific importance at the time it was published and during the subsequent 5 years. This study also found that the scientific value of review articles were largely overestimated. The relationship of importance among publications were not well reflected by the citation pattern according to this study. Among these three articles, one of them used interviews to investigate the assessment of authors while two of them used questionnaires. For the first article, interviews were mostly used to investigate authors’ perceptions of scientific progress and for peer-evaluation, which is different from the aims of the current study. But its methods of comparing the best publications with the highly cited publications suggested the initial idea of the current study. The four main dimensions of the second article provided the idea on analysis the reasons of importance of this study. The angle of the third article was about scientific contribution while this study is about importance of the publications, which is on more personal perspective. Moreover, as a qualitative analysis study, the current study paid more attention on how the external and internal factors affected the citation counts of publications.

3 Methodology

3.1 Research approaches

The questions of this study is to investigate authors’ experiences on citation counts and their perception of importance of their works and then explore the reasons to potential situation. To achieve the aim, research approaches and the reasons are listed below:

Subjective: The studies require the investigation of authors’ perception about the importance of their works and discussion to explore the reasons, which data collected from authors’ subjective feeling rather than objective records.

Inductive: The conclusion of this study are designed to come from analysis on investigation results.

Qualitative: According to the subjective data collection and inductive research process, the study is a qualitative one which focus on narrative data and participants with familiar characteristics. The authors’ experiences about the citation were investigated. The study paid attention to the process which caused the outcome such as how it happened, why it happened rather than merely focus on the relationship between the input and output. The conclusion of this study could only be applied to the groups under investigation while could be used for reference by other groups. The process of the study is hoped to produce an assumption that need to be verified.

3.2 Methods of Investigation

According to the question of this study, authors’ perception on importance of publication is planned to be collected. The most appropriate methods to collect this data is through face to face interviews. In the investigation process, participants were asked to list their ten important works before interviews through emails. During the interviews, the reasons why they think those result in the high, moderate, low or even negative correlation between their evaluation and citation counts were discussed. The type of interviews will be unstructured with opportunist sampling and small sample size. The participants of interviews will sampling academic staff of information school in University of Sheffield as they are also authors of many publications. The sampling method is determined by its accessibility and this opportunist sampling is recognized in the literature as being both practical and efficient (Robson, 1993; Bell, 1987). The similarity of research fields, research environment as well as the same institution they affiliated would reduce the influence of external factors on exploring of questions.

3.2.1 Interview design

Interviewees were asked to list their most important publications (up to ten) through emails. Then the questions of interviews were designed according to the citation counts from Google Scholar and Web of Science of these publications. The questions were mainly about the reasons why authors think these publications are important and their perceptions on the results of the correlation between the importance and citation counts. The listed publications of an author which with low citation counts as well as the publications with high citation counts but not mentioned in the list were both focused on. The interviews were recorded and all results presented were anonymous. These were fully informed to participants to ensure the obtained answers could exactly reflect real condition.

3.2.2 Interviews Preparation

The target participants were contacted through emails. In the invitation emails, a brief and clear introduction of the study and interviews including required time and possible questions were given to make sure they are fully informed. The information sheet and approval letter were attached. Next step started only after their consent of participation. With the consent of participation, participants were asked to list their ten important articles. The date and place of
the interviews were also confirmed through the communication with participants. Before interviews, interviewers have searched for the citation counts of the listed publications. Publications with high citation counts but not mentioned in the lists were also recorded and discussed during the interviews.

3.2.3 Interview guide design
The interview is designed to divide into three parts:
Firstly, making a brief introduction on study and asking interviewees to sign the consent forms.
Secondly, asking participants to give a brief introduction on each publication about why they think the publication is important; then discussing on reasons of importance to find out the main reasons.
Lastly, discussing the results of correlation between perceptions of importance and real citation counts and discovering the potential reasons.
There were some wording issues considered: asking one question at a time; making sure no stance towards in every question; giving a scope of question to make it easier to answer.
Strategy of asking: asking a more detailed question bases on pervious answers to get insight of discussion; using role play and assumption condition to make a question easier to answer and more precise on meanings.
The total time cost was about 10-15 minutes.

3.2.4 Data Recording
All interviews were recorded while brief notes were also taken about crucial points during interviews. After interviews, the record of interview was allocated into text record according to recordings, notes, and memory as soon as possible.

3.3 Results analysis
There are two steps of results. The first is the extent to which citation counts correlate with authors’ perception of the importance of a publication. The correlation and significant level were calculated to indicate the results of this stage with three different measures. This results could only be applied to the sampling group for the methods of sampling.

The second stage is to analyse the reasons which lead to the results in the first stage. The records of the interviews were deeply analyzed to conclude the potential factors. Several assumptions about how these factors effected the citation counts were made and the phenomenon behind the results was discussed in the end.

3.4 Ethics issues
This study is with low risks for human participants were used as respondents and people can make informed decisions. The topic of the research is not in any sensitive areas. For participants, the risks of participating are the same as those experienced in everyday life.

The interviewees of this study were members of the academic staff of the Information School in the University of Sheffield. They were asked to send to the details of what you consider to be their ten most important publications. Then there was a short interview to discuss the correlation between their perception of importance and the real citation counts of these publications, and to discuss the possible reasons for any obvious discrepancies. To ensure the informed consent, the consent form and informed sheet were attached to the invitation emails to participants.

In terms of data management, the details of these publications and the interview were recorded. The interview data was used to analyse to determine the overall correlation between the authors’ perceptions of importance and the research community’s perception as indicated by citation counts. After this dissertation, the data will be destroyed. The stored data was anonymised and the computer files were coding with alphanumeric identifiers. That said, given the nature of the sample, people with a knowledge of the field might be able to identify, or at least to guess, at least some of the individuals involved by doing citation searches in the Web of Science and Google Scholar. The results of this study were included in this dissertation which will be publicly available.

4 Findings
The purpose of this study is to find the extent to which citation counts correlate with authors’ perception of the importance of a publication. Furthermore, investigations also have been done for trying to explore the reasons and the phenomena behind the results.
This study obtained data from 8 academic staff among Information School from the University of Sheffield. For 7 of them, their perceptions were collected through the interviews while 1 of them answered questions via emails (for this participant was too busy to arrange time for an interview). In consideration of privacy of the authors, presentation in this dissertation is anonymous. 8 authors are coded as A1 to A8. All citation counts presented here were from Google Scholar and Web of Science and updated until 24/08/2015.

**Question 1.** The extent to which citation counts correlate with authors' perception of importance of a publication.

To investigate this question, participants were asked to list their publications (up to ten) which they think are most important through emails. These lists of publications were used to be compared with their top cited publications lists from Google Scholar and Web of Science in order to find whether there are some inconsistency between the two lists. Since many publications could not be found in Web of Science, the top cited lists used here are all from Google Scholar. For each author, same amounts of publications from top cited list were chose as the amounts of publications they listed in the important list. Totally, 76 important publications were provided by 8 authors. One of them provided 6 publications while the others provided ten. The author who provided 6 publications has published 6 publications currently.

At the first stage to answer this question, the percentages of important publications which also listed in the top cited lists are used to present the extent of the correlation. Since 2 of the participants (A5 and A6) had proved all of their publications as their important publications, the importance of publications they listed might be not that significant as others. They provided all of their publications as the important publications probably because they have only published ten or less than ten articles. The results of each authors need to be presented here to reflect the influence of this condition on the total results. A3 has three of her papers are cited three times, which is all tenth in her top cited list. In this condition, the recent one is considered as the more cited one since it was cited same times in less year as other two papers. For some book chapters, there are only citation counts of the whole books can be found and it is difficult to distinguish the citation of the whole books from the certain chapters. So the citation counts of the books are used directly in this study.

The percentages of important publications amounts in the top cited lists of total and each author are showed in figure 4.1.

![Figure 4.1 Percentage of important publications amounts in the highly cited lists](image)

As showed in Figure 4.1, the percentage of total amounts is 68.42% and the lowest correlation among percentages of each author is 40%. Which could initially indicate that the citation counts could reflect the quality of a publication to a certain degree, but there are still some other factors that might affect the citation counts. These factors are discussed in the subsequent questions.

What needs to be mentioned is that, A5 and A6 listed all their publications (6 among 6 and 10 among 10 respectively) as the important publications while A8 listed 10 among his all 11 publications. It probably because their career lives are not that long as others currently and one of them mentioned that he focuses more on teaching than academic researches. Hence the importance of the publications they listed might be
not as significant as others. This condition would increase the results of correlation so that the correlation degree might not as much as that presented in the figure 4.1.

In some cases, the citation counts of the publications at the end of the top cited lists is not that markedly higher than publications which are not in the top cited lists but mentioned in the important lists. For example, the tenth article of an author’s top cited list was cited 21 times while one of his important article was cited 20 times. Then the latter would not be listed in the top cited list. It is not reasonable to say that the latter is not as highly cited as the former just because the only once citation. However, the latter would actually be regarded as the “important article without high citation counts” according to the method above. This condition would influence on the results a lot by using the amounts of publications to present the extent of correlation.

So it comes to the second stage of results to solve this problem. In this stage, the citation counts of publications instead of the amounts of publications are used to present the extent of the correlation. At this stage of results, listed publications were classified into three parts: publications both listed in the important lists and top cited lists, publications only listed in the important lists and publications only listed in the top cited lists.

This method could involve the significance of difference of the citation counts into consideration. For example, an article with very high citation counts but not be mentioned in the important list would increase the percentage of that part greatly. What’s more, important articles which not be listed in the high cited list only because of the few citation counts difference could offset the influence of articles at the end of the high cited list relatively when comparing these two parts.

For there are many publications could not be found in Web of Science, the citation counts used here are from Google Scholar. Results from data of Web of Science are presented later in this question.

By using this method, the percentages of the total citation counts of publications in three parts are showed in Figure 4.2.

![Figure 4.2 Percentages of the total citation counts of publications in three parts](image)

As can be seen in figure 4.2, the citation counts of publications which are not regarded as important but highly cited are twice the counts of which are both important and highly cited. This results is different from results in the first stage when involving citation counts into consideration. The reasons of this condition could be explored by considering the influence of individuals. The percentages of each author are showed below (figure 4.3-4.10).
A1

- Important & highly cited: 169 (23%)
- Not important & highly cited: 556 (74%)
- Important & not highly cited: 21 (3%)
- Important & not highly cited: 1 (0.3%)

A2

- Important & highly cited: 77 (35%)
- Not important & highly cited: 137 (62%)
- Important & not highly cited: 6 (3%)
- Important & not highly cited: 1 (0.3%)

A3

- Important & highly cited: 25 (19%)
- Not important & highly cited: 101 (76%)
- Important & not highly cited: 7 (5%)
- Important & not highly cited: 0 (0%)

Legend:
- Important & highly cited
- Not important & highly cited
- Important & not highly cited
- Important & not highly cited
In figure 4.3-4.10, 5 of 8 participants’ perception of articles could be considered as highly correlating with their citation counts. 1 of 8 participants (A2) is with moderate correlation while 2 of them (A7 and A4) are with negative correlation. Among the participants, A7’s citation counts of both important and highly cited publications accounts for only a quarter of the total. Because his total citation counts is extremely higher than others, the result of total counts is effected a lot by his perception. On the one hand, it is reasonable to say that authors with higher citation counts whose perceptions are more important than others. On the other hand, using the average percentages rather than total citation counts could eliminate the influence of the differences of citation counts between authors. This method could consider all authors on the same weight. The average percentages of all authors are showed in figure 4.11.
As can be seen from figure 4.11, the percentage of important and highly cited publications is 71%. This result indicates that author's perception of a publication has moderate correlation with its citation counts. This result is similar to the result in the first stage of this question. Again, the condition that some authors (A5, A6 and A8) provided all of or almost all of their publications needs to be involved into consideration. Thus the actually degree of correlation might be a little bit lower than 71%.

In the third stage of results, citation counts per year of a publication instead of the total citation counts are used here to reflect the correlation of authors’ perceptions and the citation counts. The reason is, most of authors provided their recent published papers as their important publications and these papers are generally rarely cited. Which indicates a fact that it is unfair to compare the citation counts of recent published papers directly with the papers that have been published for many years. By using citation counts per year, the percentages of all authors and of each author in three parts are showed in the pie charts below (figure 4.12-4.20). For publications published in 2015, their publishing years were calculated as 0.5 year.
A5

- important & highly cited
- not important & highly cited
- important & not highly cited

34.16666667, 100%

A6

- important & highly cited
- not important & highly cited
- important & not highly cited

18.57142857, 100%
In terms of the percentage of important and highly cited publications of each author, the result is not much different from that in the second stage. The difference between percentage of the not important but highly
cited publications and of important and highly cited publications is decrease from 32% to 12%. Which causes by both the decrease of the percentage of the not important but highly cited publications and the increase of the important and highly cited publications.

Again, for the reasons mentioned above, the average percentages of the total citation counts of publications per year in three parts are presented in figure 4.21.

![Figure 4.21 Average percentages of the total citation counts of publications per year in three parts](image)

According to figure 4.21, the percentage of important and highly cited publication is 69.13% when using the citation counts per year to calculate the correlation. Which is not much different from 68.42% in the first stage and 71% in the second stage. But the proportion of the other two parts changes a lot in this stage. In order to track this changes intuitively, the results of total citation counts of stage 2 and stage 3 are showed in figure 4.22.

![Figure 4.22 Results of total citation counts of stage 2 and stage 3](image)

According to figure 4.22, it witnesses a great increase of proportion of important and not highly cited publications from 3% to 30.50% when using citation counts per year. Which indicates that there are some publications are not highly cited because they are recently published while some publications listed in the top cited lists only because they have been published for year thus have more chances to be cited. The influence of publishing year is discussed in later questions.

As has been mentioned, the results above are all from the data collected from Google Scholar. Since citation counts of many publications could not be found in Web of Science, it seems not convicitive to use these data.
for comparison. However, citation of Web of Science is generally regarded has higher quality than this of Google Scholar. So it is can be assumed that the results might be more accuracy by using sufficient data from Web of Science. Unfortunately, with citation counts 53 among total 105 publications involved in this study could not be found in Web of Science, the results from Web of Science might not sufficient enough to indicate the correlation. The amounts of publication could be found in Web of Science shows significant difference between authors. It suggests that, in order to obtain better results in further studies, the academic levels and career length of authors are need to keep in the same standards. On the other hand, the evolution of results among stage 2 and stage 3 could also shows the same situation of the publishing year’s influence. The average percentages of 2 stages calculated from the data of Web of Science are present in figure 4.23 and 4.24.

![Figure 4.23-4.24 Average percentages of two stages calculated from the data of Web of Science](image)

In figure 4.23 and 4.24, proportion of important and not highly cited publications also witnesses a significant increase from 2.38% to 11.13%. This result could verify the conclusion of the potential of being highly cited of the recently published and important publications.

In conclusion of the answers to the first question, the correlation of authors’ perception with the citation counts is 68.42% when simply considering the amounts of the publications. But this method seems not convincing. When introducing citation counts instead of the amounts, things are different. However, the results in this stage are easily been effected by perception of authors who have extremely higher citation counts than others. This influence could be eliminated by using average percentage of each authors. The
average percentage is 71%, which could sensibly indicate that, in this study, the authors’ perception of importance of a publication has moderate correlation with the citation counts of this publication. In the third stage, when taking the publishing year into account, the average percentage of important and highly cited is 69.13%, which further verifies that the results in the previous two stages. Moreover, the great increase of proportion of important but not highly cited publications in this stage indicates the potential of being highly cited of the recent published publications. As this study failed to obtain enough data from Web of Science to indicate the correlation, the evolution of proportion of important and highly cited publications also indicates the potential of being highly cited of the recently published publications in important lists.

In terms of individuals, most authors’ perceptions are not that different from the citation counts while there are some exceptions. Since authors might have different preference when choosing their important publications, it is necessary to realize the reasons why they regard the publications listed as important.

**Question 2. Reasons of publications are regarded as important.**

In consideration of the fact that authors might be effected by some factors not relate to the content of articles when choosing their important publications, the reasons why they regarded the publications listed as important are necessary to be realized. The factors which not relate to the content of publications is called “external factors”. Some of the external factors (especially citation counts) would affect the results of correlation a lot.

In order to realize the effect of the external factors, participants were asked to give a brief introduction of each publication about why they think it is important during the interviews. According to the answers to this question, the factors are divided into 10 categories below and some fragments of the interviews are presented as examples of their related categories. Among these categories, three external factors are identified (categories 8-10) because they have no direct relation to the quality of publications.

1. **Topic and methodology.** Some authors regard their publication as an important one because they think its topic is interesting, the idea is clever or its methodology is good.

   “...it based on quiet lot of data and I quiet like the idea those developing in that.” [A1]

   “And it was a really fun thing to do.” [A7]

   “All of these three...which showed something that was broadly accepted, was wrong...I was really proud of that, I think it is a really clever thing.” [A7]

2. **Research area.** Some authors regard their publication as an important one because the research areas of the publication are highly concerned by people or the publication is the pioneer one in its research area.

   “...but then use MCDA in e-procurement...a combination of this study has not been explored that well.” [A2]

   “So again it is the hot concept in the IT industry.” [A2]

3. **Type of publication.** Most authors think that theoretical article is more important than a practical one while some of them believe publications about methodology are relatively important.

   “...articles about research methodology are always important from the theoretical perspective.” [A2]

4. **Valuable content.** Some authors regard their publication as an important one because its content is valuable. For example, it develops a theory or acts as the basis of a further study.

   “...so there will be some interesting practical values...So it have both theoretical and practical values.” [A2]

   “...and I am including them for their originality and contribution to the field.” [A3]

   “...because it involves some key arguments...in terms of developing my theoretical understanding.” [A5]
5. **Potential application.** Some authors regard their publication as an important one because its potential application or it could bring potential benefits.

   “...so there will be very important practical benefits and practical implications behind this study.” [A2]

6. **Author’s Effort.** Some authors regard their publication as an important one because they have done a lot of efforts on it.

   “Again I think it is a quiet nice paper and I did a lot of work on it.”[A1]

7. **Positive feedback.** Some authors regard their publication as an important one because they have got some positive feedback of this publication.

   “This one was dominated for...so it is in top 2% papers of this conference and that is why I put it among my best publications.”[A4]

8. **Publishing issue.** This category of factors is regarded as external factors. There are normally two kinds of publishing issues, publishing place and publishing time. Publishing place was mentioned frequently as the reason of importance, especially the level of journals and conferences.

   “This one...may be just journal is important, because it is JASIST. I don’t think actually it is a good article but the journal is JASIST.”[A1]

   “This one...the topic is different from others but it is also from this very good conference.”[A4]

9. **Career life and personal experience.** This category of factors is regarded as external factors. This kind of reasons is usually mentioned on the condition that, the publication is author’s first one has been published, the author acted as the only author of a publication or the publication helped author to build reputation in a certain research area.

   “The contribution of these three papers that actually helped me to establish the reputation in the ERP field.”[A2]

   “One of my first published papers, and one which involved an enjoyable collaboration with 2 people whose work I respected.”[A3]

   “This was the first paper I wrote to be awarded a prize...so for this reason I see it as an important milestone in my research career.”[A3]

10. **Citation counts.** This category of factors is regarded as external factors. Some of authors mentioned that publications with high citation counts could bring them positive feedback or make them feel their researches are valuable.

    “This is been cited quiet lot... I am not sure it is that good but it has been cited a lot. I can’t really ignore that, can I? ...I wish that people found it useful...” [A1]

Actually, citation counts could be considered as a kind of positive feedback (category 7). But in this study, they are regarded as different categories since positive feedback could sometimes reflect the quality of a publication directly while citation counts is a factor which would influence the results of this study that needs to be distinguished from other factors.

Then next step is to analyse which reasons that mainly affect the perceptions of authors of importance by using the frequency of those reasons were mentioned. For an article, the author may give more than one reasons, and all of these reasons are involved in the analysis of this step. And for the situation that for one article, the author may mention two reasons which belong to the same category, then this category is calculated twice. For reasons that could be involved into two categories, they were involved into both categories. For example, reason “it described a project which of all the thing I have done for 17 years” [A7] is involved into both “career life” and “valuable content” categories. It is also need to emphasize that there could be some misunderstands of perceptions of authors in spite of researcher tried to avoiding that. But it is hard to be eliminated totally.
By using the methods above, the frequency of each category of reasons are summarized and presented in the column graph figure 4.12.

![Figure 4.12](image)

**Figure 4.25 Frequency of each category of reasons on importance are mentioned**

According to figure 4.12, most reasons provided could reflect the quality of papers to a certain degree (categories 1-7) while some external factors have considerable influence. Among reasons 8-10, citation counts was only mentioned twice within total 77 results. Which indicates that authors’ perceptions in this study are with few influence by the citation counts of publications. However, categories 8 and 9 show high frequency of occurrence. While published journal of conference (category 8) could sometimes reflect the quality of publications in a certain degree, the reasons about career life (category 9) seem have no relation to the quality of publications. The proportions of these two categories are 15.58% and 19.48% respectively. Which indicates that the effect of these external factors need to be involved into consideration when calculating the degree of correlation between the authors’ perceptions of importance and citation counts. Hence the real correlation is lower than the results in the first question.

In this stage of study, authors were found to tend to have preference when stated the reasons of importance. For example, A3 presented 7 of her 8 reasons relating to her career life including helping to building the reputation in a certain area, publishing at first time, being awarded at first time and so on. As for other authors, the situation is not that extreme but it also shows a tendency that young authors prefer choosing the articles which may affect their career lives while experienced authors more focused on the topic and content. There is also a difference that some authors preferred choosing their important papers mostly because they are “interesting” or “topic is clever” while some of authors more focused on the potential applications and benefits of their publications.

On the other hand, there is another factor about preference that may influence the results. That is, when authors were stating why they think a publication is important, they might give the reasons which firstly emerged in their brains rather than all reasons because of the limitation of the time of interviews. For example, when an author gave a reason “interesting topic” that not means he or she has not done a lot of efforts on it, it just his or her preference make the former as the mainly reason. To avoid this problem, in further studies, options are suggested to be provided to authors to select reasons which fit their minds. Which could also reduce the misunderstandings in a great extent.

In conclusion of this question, external factors which have no relation to the quality of publications such as publishing issues and career life that may affect the results in the first stage. The proportions of these two categories among the total reasons are 15.58% and 19.48% respectively. Which needs to be involved into consideration when judging the extent in the first question. However, it needs to mention that there is many
limitations of the analysis of the reasons on importance in this question. So it could only be regarded as a rough direction rather than a precise result.

**Question 3. Authors’ perception about why their important publications are not highly cited.**

To answering this question, authors need to give some speculations about the reasons why their important publications are not highly cited. The reasons and the frequency they were mentioned are presented in the chart figure 4.25

![Figure 4.25 Reasons of important publications are not highly cited](image-url)

From figure 4.25, it is easy to see that publishing time has the most frequency of occurrence. Many publications were never cited or few cited because they were recently published. This condition may influence the results a lot and it has been take into account at the third stage of the first question. The reason of research area is, for some topic, people are not interesting in them currently. For the reason of publishing place, authors regarded that the publications were not highly cited because of the effect of the place they were published are not good. For example, the journal is not good so that it may attract less people to read and to cite the publications. Among the reasons of accessibility, there are two kinds of condition of accessibility. The first one is not easily accessible to read for the publication is a book chapter or only accessible on a certain website. The second one is because the content not accessible to non-academic people for the article is too theoretical.

The result of this part is mainly about the external factors that may have negative effect on the citation counts. It is necessary to take them into consideration when using citation counts to evaluate a publication, an author or an academic group.

**Question 4. Reasons of highly citied publications were not regarded as important and reasons why they were cited.**

With the same aims of question 2, the top citied lists were showed to authors and the publications which were highly citied but not listed in the important lists were highlighted. Authors were asked that why they think these publications are not as important as others. For authors who provided all of their publications as important publications, this question was not asked. The answers to this question and the frequency they were mentioned are presented in the figure 4.24.
<table>
<thead>
<tr>
<th>Category of reasons</th>
<th>Reasons</th>
<th>Frequency</th>
<th>Mentioned by whom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publishing issues</td>
<td>The publication is very old</td>
<td>3</td>
<td>A1, A1, A1</td>
</tr>
<tr>
<td></td>
<td>Conference paper</td>
<td>1</td>
<td>A2</td>
</tr>
<tr>
<td></td>
<td>Workshop of a conference</td>
<td>1</td>
<td>A4</td>
</tr>
<tr>
<td>Type of Publications</td>
<td>Good but quiet practical</td>
<td>1</td>
<td>A1</td>
</tr>
<tr>
<td></td>
<td>Review article</td>
<td>2</td>
<td>A7, A7</td>
</tr>
<tr>
<td></td>
<td>book, collection of papers</td>
<td>1</td>
<td>A7</td>
</tr>
<tr>
<td>Content</td>
<td>A simple description of data</td>
<td>1</td>
<td>A4</td>
</tr>
<tr>
<td></td>
<td>Not that good</td>
<td>1</td>
<td>A7</td>
</tr>
<tr>
<td></td>
<td>An original idea</td>
<td>1</td>
<td>A4</td>
</tr>
<tr>
<td>Other reasons</td>
<td>Not as main author</td>
<td>1</td>
<td>A1</td>
</tr>
<tr>
<td></td>
<td>Act as advisor, not author</td>
<td>1</td>
<td>A7</td>
</tr>
</tbody>
</table>

Figure 4.27 Reasons of highly cited publications were not regarded as important

These answers are not only the reasons why the publications are not important but also refer to the reasons why these publications were highly cited. For example, when authors mentioned that an article is very old, it probably not because they think this article is not important because it is old and it may because it is highly cited because it is old. The fragments of each reason are presented below for further discussion. In presentation here, publications mentioned are coded from P1 to P11 in order to distinguish the condition that some of these reasons are describing the same publications and they might different on importance.

“That one is very old, and the reasons it was cited might be it was old…it got more time to been cited, isn’t it?…” [A1, P1]

“This is also old…I think it is not probably quiet good but it is very practical…” [A1, P2]

“This is from my master dissertation…1996…so it is definitely old.” [A1, P3]

A1 had mentioned a fact that the old publications have more chances to been cited, so they are highly cited publications not because they are that good and just because they are old. Which indicates the importance of using the citation counts per year rather than the total citation counts in discussion of extent of correlation.

“As you can see here it is a conference paper… I would not personally say that conference paper is not important but then I think for the UK education system nowadays, conference papers would not be counted as important…even though the citation counts is good, it seems that this paper is attracting some good attention…but when we are thinking about our career in UK, thinking about the system here…other colleagues might not think it is important because it is a conference paper…” [A2, P4]

“This paper was published in a workshop of a conference, so it is even less important than the conference, and that is why I didn’t put it in the list…” [A4, P5]

These are reasons about publishing place. There shows difference that in further description of their papers, A2 regarded this paper as a valuable one which is about a research methodology while A4 didn’t think the paper itself is that good. Which indicates that A2 considered the publishing issues as a mainly reasons and A4 was not. Moreover, A2 had put forward an interesting point, that is, in UK academic system, conference papers are generally not regarded as important as journal papers. And this point will be discussed in the discussion chapter.

“This one I didn’t count it because it is quiet practical…” [A1, P6]

“It is very similar in basic concept to that one there (an article that was regarded as important), so that is why I didn’t include it. Basically…I have too many good papers.”[A7, P7]
These reasons are about the content or about the type of the publications. These two kinds of reasons are sometimes overlapping so they are presented together here. For the two reasons above, they are reasons which make authors feel the publications are not important.

“These are two most cited review articles that I have ever written, and when I wrote them, I think they were very good piece of work, and also been very extensively cited...most other people thought they were important as well. However, they are reviews rather than original work...so that is why I haven’t including them as most important paper since they are not basically researches, they are review articles.” [A7, P8]

“…That is a book...there is a series of carefully selected papers, which we thought were most important papers in information retrieval. So it is a collection of papers and was highly cited...I wouldn’t include that one as my important paper because basically it is a collection...It highly cited in the same way of those review articles are highly cited...It contains a huge amount of very useful information. But it is not one of my intellectual contribution to the discipline.” [A7, P9]

“This paper is a description of data set we produced, and usually everyone who uses these data cited it...but the paper is just a simple description of data... and it is not like research one, is not high quality research...and it received many citation just because who uses this data set will cited it.” [A4, P10]

“…also this paper is nearly a primary work with fewer experiments, but it was an original idea, that is why it got many citations. But this paper itself is not very good...it probably because many people work on it and cited it. But this paper itself is not a fantastic one.”[A4, P5]

These reasons indicate why authors think their publications were highly cited on the views of content and the type of the publications. According to these reasons, review articles (P8) as well as the collection of papers (P9) are easily be cited for there are many useful information in them. Analogously, the paper which describes a set of data (P10) also highly cited for the same reason. A4 also mentioned that publications about original idea are tend to get many citation. The influence of the type of publications is discussed in the discussion chapter.

“…and also the main author was a student, I didn’t do that much on it…” [A1, P6]

“This one here… there are three other people...who are advisors. And my name is just be added in to the authors’ list...I just an advisor to this paper.”[A7, P11]

For these reasons above, authors not included them because actually they had not done much on them. Also the condition of homonymies was mentioned by A7 during the interview of this question. Fortunately, that article is not included in the highly cited list so that it not effected the results of this study. But it is still a condition need to be awarded of in further studies of citation analysis.

“…We then work on it for two years… and then published the second paper, which is this one. And historically, people always cited the second one rather than the first one. I don’t know why, in my view that is more important one since that was the thing said why we invent this wonderful new way of doing something...in the second paper, we showed the way in which you could test this sort of algorithm, and we provided a test set of compounds which a lot of other people used subsequently. So in my view that is the more important paper since that described the basic idea (the first paper), and that is the extension of the idea (the second paper)... and for whatever reason people cited that much more than cited that.”[A4, P12]

Reasons for this publication are not included in the table above because the author did not gave a clear reason. In this case, people cited the extension of the idea rather than the basic idea.

In conclusion of this part, there are many different reasons for authors did not regarded some of their highly cited papers as important papers and reasons why these papers are highly cited. Specifically, some old publications have more chances to be cited so they are highly cited not only because they are good.
Moreover, some publications are highly cited for they contain much useful information, which is not directly related to the quality of them.

**Question 5. Inconsistency between importance and citation counts among the important publications.**

Since the citation counts of the publications in an important list are normally different, there should also be inconsistency between importance and citation counts among the important publications. For example, an author might find his or her most cited article is not that important or the citation counts of an article is far from his or her expectation.

The conditions of this kind of inconsistency are presented below.

“...I am not sure it is that good but it has been cited a lot.”[A1]

“...is my most cited article, it is definitely not the best article...there are more theoretical content in these two articles (two better articles).”[A5]

The condition of this kind of inconsistency is not very common among authors in this study. So it might not has significant influence on the results of correlation. Also some authors reported that, for some important publications, the citation counts is lower than their expectation. This condition are not take into account in this study for it may not have influence on the results of the first question.

**5 Discussion**

In findings chapter, there are three stages of results of the question “to what extent the citation counts of a publication correlate with author’s perception of importance of it”. All three stages of results indicate that the correlation is moderate since the both important and highly cited publications occupied 68.42%, 71% and 69.13% among the total publications respectively. Since the aims of this study are not only to define the extent of correlation but also to explore the reasons and phenomenon behind the results, it is necessary to have further discussion on which factors that may affect results and how these factors exerted the influence on the results. With the results of this question being similar among three different measures, the evolution of the results might reveal the influence of some factors.

Results of the second stage came out by taking the significance of difference between citation counts into account. However, the results of the second stage are not that different from those in the first stage. Which could indicate that the condition of the significant difference is not common enough to effect the results. More in detail, this condition was only occurred in A7’s result that his top 2 highly cited papers were not regarded as important by him. The citation counts of these two papers is 3552 and 1336 respectively while the third paper got 1143 citation. The significant high citation counts actually has effected the results of total citation counts in second as well as the extremely higher citation counts of A7’s publications than other author’s. But the influence is not obvious when considering all authors on a same weight by using average percentage. Furthermore, on considering the fact that some of author’s publications could not be found or have no citation on Web of Science, the length of career life and academic level of an author are probably need to be considered when processing an analogous study.

The remarkable changes between the results of stage 2 and stage 3 reveal the influence of publishing year. The significant increase of proportion of important but not highly cited publications indicates that there are many recent published papers were included into the important lists. To investigate the influence of this condition, the proportion of important but not highly cited publications in stage 2 and stage 3 as well as the amount of publications publishing in recent 5 years (including 2010) of each author are present in the figure 4.25.
As can be seen from figure 4.28, the amounts of publications publishing in recent 5 years is really impressive. This condition again indicates the importance of taking the publish year into consideration when processing citation analysis. Also the amount of publications publishing in recent 2 years are listed here on considering the phenomenon that a scientist is expected excited by what he is now doing (Porter, 1988). In additional, the length of career life of authors also need to be considered. For authors who have engaged in the academic research no more than five years, it is so normal that they provided all of their publications publishing in recent 5 years.

Besides publishing year, the influence of publishing place and type of publication also need to be emphasized according the results of interviews.

In terms of publishing place, A2 has put forward the condition that the conference papers are normally not regarded as important as journal articles in UK academic system. However, when discussing this problem, A4, who provided 7 conference papers as his important publications, said that some conferences may be better than journals and it depends on disciplines. These papers of A4 are from some top conferences such as International Conference on World Wide Web and ACM SIGIR Conference. That is one of the reasons why he included them as important papers. The condition are also confirmed by A7.

A2 also pointed out an interesting condition when discussing two of his good papers were not highly cited. 

“...for some SCI papers, they also tried to cite SCI papers as well to try to show that the papers they cited are all in SCI...Because unfortunately even for myself, when I am writing a SCI paper I may need to do this as well. It not because that I want to do it, but because the publishing system actually make me to do it...When I target on a good journal, a SCI top journal, further to the quality of my paper...they (editors) look at your references then they may ask you that ‘some of the articles that you cited are not SCI, these journal may not be in good quality, why should you cite them?’ If you cited some non-SCI papers, which may not be in good quality, then your own may not be in good quality either. I disagree with this, but this is what has been used in practical terms. And therefore...for the authors to satisfied for the reviewers, the editors, then they try to cite the other SCI papers as well. But sometimes those papers are not good at all.”[A2]

He pointed out the phenomenon that some SCI papers or papers target on SCI journals prefer citing all SCI papers to show their good quality. So that the non-SCI papers are tend to be less cited. These condition needs to be considered both when working on the influence of external factors on citation counts and when analyzing citation behavior.

The factors of type of publications also need to be emphasized. Aksnes (2006) has concluded that the scientific value of review articles was largely overestimated. Results of this study also indicate the same problem. According to A7, three of his highly cited articles are review articles. He pointed out that these
review articles as well as one of his highly cited book, which is a collection of important papers, were highly cited because there are a huge amount of useful information in them. The two most highly cited review articles were very important when his wrote them. However, these two articles were written in 1998 and 1988 respectively and they are not up to date nowadays. But they still got 20 or 30 citation a year. And he thought this phenomenon also shows how lazy people are. People might tend to cite what has been highly cited regardless of whether there are some better choices. Similarly, as has been stated in question 4 of findings chapter, a description of data set also tend to be highly cited. Furthermore, when discussing why two of her best articles were less cited than the most cited one, A5 has stated that the latter is highly cited because it is more accessible to non-academic people while the former are too theoretical. However, she regarded the theoretical papers as more important than others.

As there are so many different reasons provided by authors on choosing their important publications, there comes a question that what is the “important” means? And what is difference between the regraded important publications and the highly cited publications? Different from the original assumption of this study, the reasons of the important are not always relate to the quality of the publication. The reasons of “important” are likely to be a more personal things, which could relate to authors’ career life, their experience or their own feeling about the topics and ideas, regardless of the perception of others. According to Aksnes’s (2006) study about the correlation between citation counts and authors’ perception of contribution, generally, citation counts have the highest accuracy in identifying either major or minor contributions. The correlation between citation counts and perception of contribution was much higher than that between citation counts and perception of importance. Besides the influence of the measure of investigation, reason of this difference might be, the meanings of “importance” is much more complex than “contribution”. An article could be important to its research area and also could be important to authors themselves. As for citation counts, it could more reflect the scientific contribution of a publication than its quality or its special meanings to its authors. And that might be the mainly reason cause the difference of highly cited lists and important lists.

6 Conclusion

In this study, the extent to which author’s perception of a publication correlate with its real citation counts was calculated from the comparison with the publications of an author which were regarded as important with those were highly cited. With using of three stages of measures which could involve the significance of citation counts, could reduce the influence of magnitude of citation counts of individuals or publishing year of each publication, the results of total data are similarly indicate that the correlation is moderate. And for individuals, most of authors showed the moderate correlation while two of them are with negative correlation.

This study also explored the processes how some external factors effected the citation counts and the phenomenon behind the results. The remarkable changes of proportion of important but not highly cited publications indicates the great influence of publishing year of these publication. More in details, some recently published papers were regarded as important but did not have enough time to be highly cited while some old papers had more chances to be cited but they are not that important.

In addition, publishing place of a publication is also an influence factor on citation counts. According to the statements of some participants, conference papers are usually not regarded as important as journal articles excepting some top conferences. Meanwhile, SCI papers may have preference to cite all SCI papers to ensure the reviewers or editors to think them in good quality. These conditions above showed the influence of current environment of publishing system and academic system in UK.

Some publications including a huge amounts of useful information were reported obtaining high citation counts but not that important, such as review article, a description of a data set and a collection of important papers. Furthermore, the discussion about some review articles indicates the condition that people may prefer citing a highly cited paper than finding some better choices.

There reported so many different reasons for the publications why they were regarded as important. Among which, the factor of citation counts has not much influence on the results while some personal factors have significant influence. As the citation counts might more relate to the contribution and utility of a publication, the concept of “importance” may contains some personal meanings. People may regard their publication as important not only because it is useful but also because it is are meaningful to themselves or they are simply
think it is interesting. That might be the main reason causes the authors’ perceptions of importance are not highly correlating with their citation counts.

There are many limitations of this study. Since the sample of this study is scarce, it is hard to get further results. And the results could on reflect the reality of participants in this study. Furthermore, according to many facts during the studies, the different academic levels and the length of career life of authors also need to be involved into consideration. To solve the problems of this condition, it is suggested to select authors with similar career time and citation counts or to involve these factors into account in the further studies.
Reference


Lokker, C., Mckibbon, K. A., Mckinlay, R. J., Wilczynski, N. L. & Haynes, R. B. (2008) Prediction of Citation Counts for Clinical Articles at Two Years Using Data Available within Three Weeks of Publication: Retrospective Cohort Study. British Medical


Appendix 1
In this appendix lists the publications provided by authors as their important publication.


Appendix 2: Consent Form

| The University of Sheffield. Information School | A study of the extent to which citation counts correlate with authors’ perception of the importance of a publication |

Researchers
Zhen Fan, fz921217@gmail.com

Purpose of the research
The aim of this study is to examine whether author’s perceptions of the importance of a research article are reflected in the number of citations to it in the Web of Science and Google Scholar databases. Data will be collected through face to face interviews with authors and compared with citation counts to investigate the relationship, if any, between citation counts and authors’ perceptions.

Who will be participating?
The interviewees will be members of the academic staff of the Information School in the University of Sheffield.

What will you be asked to do?
You will be asked to send to e details of what you consider to be your ten most important publications. There will then be a short interview to discuss the correlation between your perception of importance and the real citation counts of these publications, and to discuss the possible reasons for any obvious discrepancies.

What are the potential risks of participating?
The risks of participating are the same as those experienced in everyday life.

What data will we collect?
The details of your publications and the interview, which will be recorded.

What will we do with the data?
The interview data collected from you and from other participants will be analysed to determine the overall correlation, if any, between the authors’ perceptions of importance and the research community’s perception as indicated by citation counts. The results of the analyses will be included in my masters dissertation, after which the data will be destroyed.

Will my participation be confidential?
We shall anonymise the stored data and coding the computer files with alphanumeric identifiers. That said, given the nature of the sample, people with a knowledge of the field might be able to identify, or at least to guess, at least some of the individuals involved by doing citation searches in the Web of Science and Google.

What will happen to the results of the research project?
The results of this study will be included in my master’s dissertation which will be publicly available.

I confirm that I have read and understand the description of the research project, and that I have had an opportunity to ask questions about the project.
I understand that my participation is voluntary and that I am free to withdraw at any time without any negative consequences.

I understand that I may decline to answer any particular question or questions, or to do any of the activities. If I stop participating at all time, all of my data will be purged.

I understand that my responses will be kept strictly confidential, that my name or identity will not be linked to any research materials, and that I will not be identified or directly identifiable (but see above) in any report or reports that result from the research.

I give permission for the research team members to have access to my anonymised responses.

I give permission for the research team to re-use my data for future research as specified above.

I agree to take part in the research project as described above.

__________________________________________________________________________
Participant Name (Please print)   Participant Signature

__________________________________________________________________________
Researcher Name (Please print)   Researcher Signature

__________________________________________________________________________
Date

Note: If you have any difficulties with, or wish to voice concern about, any aspect of your participation in this study, please contact Dr. Angela Lin, Research Ethics Coordinator, Information School, The University of Sheffield (ischool_ethics@sheffield.ac.uk), or to the University Registrar and Secretary.
Appendix 3: Approval Letter

Downloaded: 23/07/2015 Approved: 13/07/2015
Zhen Fan Registration number: 140136395 Information School
Programme: Information Systems

Dear Zhen

PROJECT TITLE: A study of the extent to which citation counts correlate with authors perception of the importance of a publication APPLICATION: Reference Number 005588

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 13/07/2015 the above-named project was approved on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

University research ethics application form 005588 (dated 09/07/2015). Participant consent form 1010488 version 3 (08/07/2015).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely
Matt Jones Ethics Administrator Information School