

Features of Academic Misconducts on Master Dissertation from a Perspective of Fuzzy Comprehensive Evaluation

Yalou Liu^{1,*}, Miaomiao Jing¹, and Junyong Han²

¹Hebei United University, Tangshan, China; ²Beijing International Studies University, Beijing, China

Abstract: As an important segment of postgraduate education, the quality of the dissertation, therefore, turns into a vital means to measure the normalization, scientificity and rigor of each college. Misconducts behavior among the master dissertation writing has long been highly concerned by the society and degree-conferring units. Hebei United University (HUU) involves 19 first level disciplines and 6 secondary disciplines for academic master-accredited fields as well as 7 categories for vocational master-accredited fields. There are over one thousand people apply for a master degree here for recent three years, which all these make HUU a key unit for postgraduate students training in Hebei province. Same as all other degree-conferring units, HUU entrusts assigned person to detect the replication and reference rate of each paper by utilizing the TMLC on the basis of China Academic Journal Network Publishing Database. The related staff found that the replication and reference rate can objectively show the basic characteristics of the plagiarism in master dissertation. This paper focus on 1234 master dissertation applying for master degrees in HUU in 2014 and do the statistical analysis of the initial replication and reference rate detection conclusion. The statistics complies according to the subject (natural science and social science), degree type (Full-time academic, Full-time vocational, professional education personnel and equivalent). The result divided into three effective intervals: replication ratio 0-19.9% (Qualified), 20-49.9% (to be modify), over 50% (disqualification). With the data we can build a mathematical model and conduct the relevant analysis and evaluation thus to obtain the essential features of academic misconducts among master dissertation.

Keywords: Academic misconducts, features analysis, fuzzy comprehensive evaluation, grey relevancy.

1. INTRODUCTION

Misconducts behavior among the master dissertation writing has long been highly concerned by the society and degree-conferring units. Both the *Instruction on strengthening the construction of academic morality and norm from the State Council Academic Degree Committee* and the recently issued *Disposal measures of falsify in academic papers* by Ministry of Education have put forward the related norms, the disposal measures also clear the responsibilities for degree-conferring units, tutors and graduates. On prevention of the academic misconducts in dissertation, degree-conferring units shall detect the replication and reference rate in papers by utilizing the system TMLC. Most units request the replication ratio below 20%, which restrain the replication ill phenomenon to some extent [1]. But few units would do the statistic analysis of the papers for the reasons of the limited time and sources and so on, thus what we got may only a conclusion rather than a combination with the quality of graduates education after the degree conferring. It's a failure of extend and reflect the function of the degree management department. On the bases of above issues, this research focus on 1234 master dissertation applying for master degrees in HUU in 2014 and do the statistical analysis of the first replication and reference rate detection conclusion. Then try to conclude the characteristics and analyze the results, so as to provide suggests for academic ethic education, innovation of

talent cultivation model and tutors performance assessment mechanism.

Plagiarism has always been a wrongdoing in both foreign and domestic academia and the most prominent problem in master dissertation writing. Article 46 in our Copyright Law define that copying others production means to use or publish other people's works without the authorization of the copyright owner and plagiarism refers to claim ownership of others works by changing the appearance without giving references, which leads to a misunderstanding that the production is out of his own innovation. Specifically, copy and plagiarism means occupying others productions by cut and patched up the originals and publish the works as their own results [2]. The plagiarism in paper writing can be various, namely: direct plagiarism (plagiarize a whole article); indirect plagiarism (pick and mix, part rewrite, plagiarize mainstream view, etc) [3].

2. MAIN CONTENT AND ISSUES OF THE RESEARCH

This paper focus on 1234 master dissertation applying for master degrees in HUU in 2014 and do the statistical analysis of the initial replication and reference rate detection conclusion. This paper is supposed to solve the following issues:

2.1. General Distribution States of the Replication and Reference Rate

The initial replication and reference rate detection conclusion, see Table 1.

Table 1. The initial replication and reference rate detection conclusion of the 1234 master dissertation in our college units %.

Plagiarism Degree	0	0 ≤ 10	11 ~ 20	21 ~ 30	31 ~ 40	41 ~ 50	>50
ratio	0.47	54.98	28.43	10.28	3.50	1.95	0.39

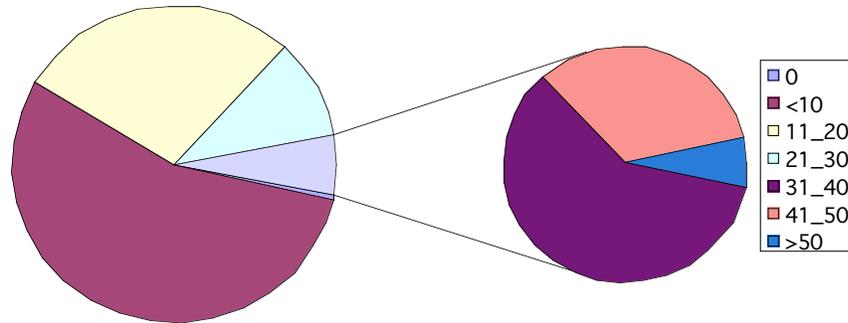


Fig. (1). The initial replication and reference rate detection conclusion of the 1234 master dissertation in our college.

Table 2. The distribution of the initial detection of the replication and reference rate among different types units%.

Replication and Reference Rate	<10	10 ~ 20	20 ~ 50	>50
Student Type				
full-time students	64.23	26.24	9.39	0.14
equivalent	44.74	33.33	20.72	1.20
on-the-job graduates	39.39	31.17	29.44	0.00

We can see from Fig. (1) that about 83.88% of the papers own a replication and reference rate below 20%; currently our college request the replication ratio need to below 20%, which means 80% of the detected papers can meet the basic requirements; the replication ratio less than or equal to 10% is a reasonable bond stipulated by related government departments; and 55.45% of the papers with a replication ratio that below 10%, in other words, over half of our papers meet this requirement [4]. Mild and moderate coincidence (replication ratio between 20% and 50%) accounts for 15.73%; severe coincidence takes only less than 0.4%, which notes that part of the detected papers have obvious plagiarism issues.

2.2. Specific Characteristics of the Plagiarism in Master Dissertation

2.2.1. Replication and Reference Rate of the on-the-job Graduate Students are Relatively High

This paper classifies the 1234 master dissertation according to the learning methods of the graduates: 690 papers from full-time students, 313 papers from applicants with education background equivalent to college graduates for master's degree and 211 papers from on the job graduates. We calculate the replication and reference rate distribution state in Table 2.

Table 2 shows the distribution of the replication and reference rate below 10% in the initial detection. Respectively, the full-time students, the equivalent graduates and the on-the-job graduates take a percent of 64.23%, 44.74% and

39.39%; we can clearly see that full-time students occupies a majority part. Besides, among the replication and reference rate that below 20%, the percentage of the above three are 90.47%, 78% and 70.56%; full-time students wins again. So we draw a conclusion that full-time students own a higher standard degree and the replication and reference rate of on-the-job graduates is obviously higher.

2.2.2. Social Science Papers Accounts for a Higher Percentage in High Replication and Reference Rate Zone

This paper divides the 1234 master dissertation into two subjects: social science and nature science. According to the above classification, there are altogether 639 sample papers among which 471 papers are natural science and 168 papers are social science. Table 3 shows the changing condition of the replication and reference rate among them.

From Fig. (3) we can see that the replication and reference rate of more than half of both social and nature science papers below 10%; over 60% papers owns a replication and reference rate that below 20%, which is on account of our college detection regulation. The replication and reference rate of social science papers is 27.71% higher than nature science papers in high rate sectors. By comparing the rate distribution among the three sectors, we got the scaling relation of the papers from low sectors to high sectors tend to be 5:2:1 and 2.3:1:1.6. The distribution difference may be related to the distinction of the two subjects. Then we come to a conclusion: humanities and social science papers may depend more on literatures than nature science so it's more likely to plagiarize. On the contrast, nature science needs

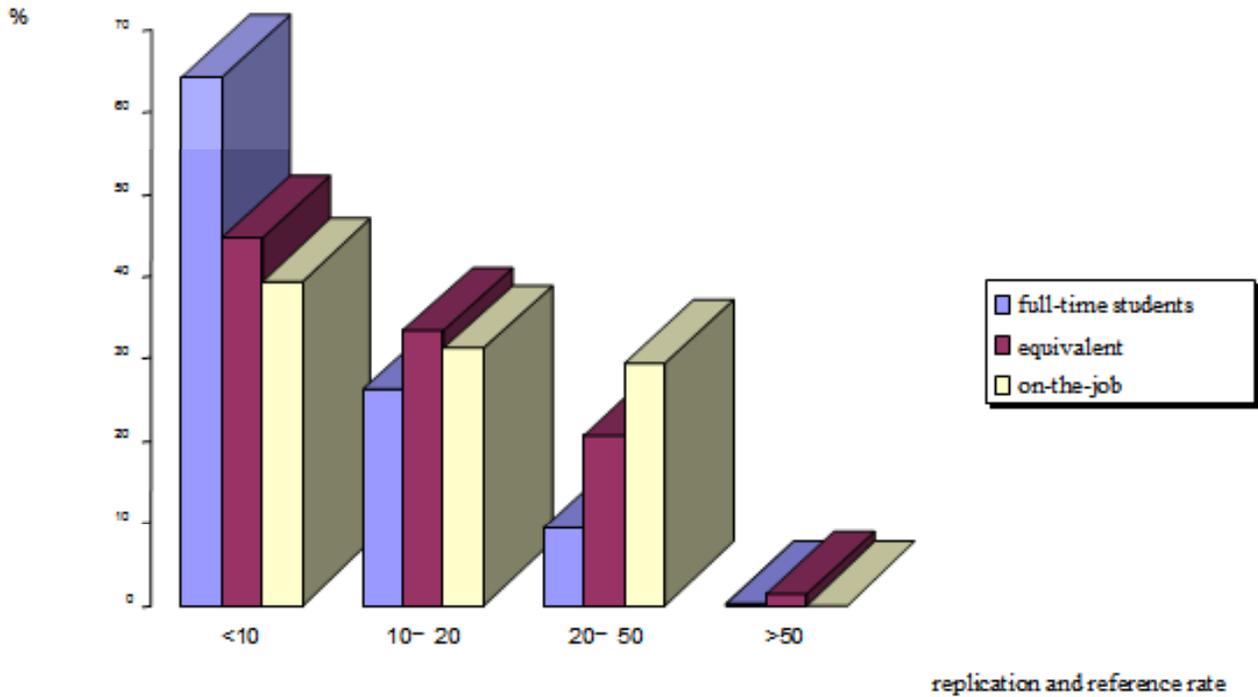


Fig. (2). The distribution of the initial detection of the replication and reference rate among different types.

Table 3. Replication and reference rate of the two subjects units %.

Subjects \ Rate	<10	10 ~ 20	>20
nature	62.65	24.70	12.65
social	57.02	25.11	39.36



Fig. (3). The initial detection of the replication and reference rate of the two subjects.

more practical operations or experiment while finishing the papers, thus to objectively reduce the chance of plagiarize.

2.3. Analyze the Sources of Replication and Reference Literature

In this research, we mainly analyze the literature sources emerged in high rate of replication and reference papers and conclude the categories and amounts of the literature [5]. We select papers with a high rate of replication and reference so as to get rid of the effect of subjective will, thus to study the dimensional representation of plagiarism.

Table 4 shows the distribution of the quantity and sources of the literature among the 1234 papers of our college in 2014.

Through the analysis, we find in high rate sectors there are three major sources including: master and doctoral dissertation, academic journals. Among the 1234 papers, 70.59% of the papers plagiarize less than ten literatures, 14.56% of the papers plagiarize about 10 to 20 literatures; only less than ten papers plagiarize 30 literatures. Furthermore, certain papers plagiarize nearly 40 literatures; very few of students have not plagiarized any of others works.

Table 4. Distribution of the quantity and sources of the literature units: %.

Scope \ Ratio	<10	10 ~ 20	≥ 20
Selected social science papers in same sectors	54.13	28.27	17.6
Selected nature science papers in different sectors	53.37	32.53	14.10

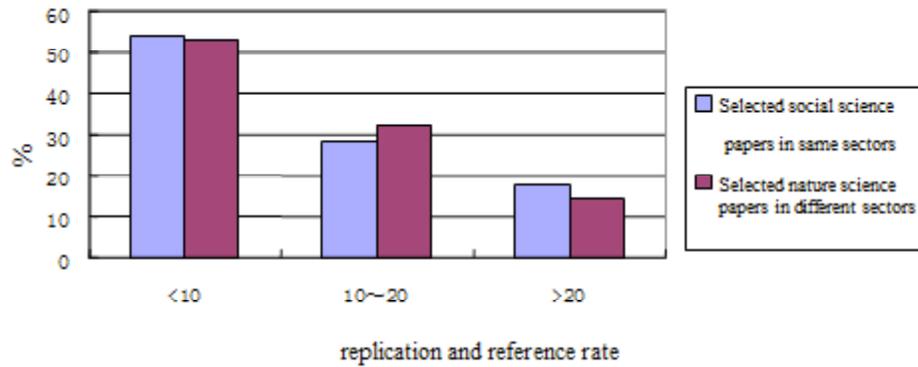


Fig. (4). Distribution of the quantity and sources of the literature.

Table 5. Distribution state of the sources of the plagiarized literatures and the distribution of the number of plagiarism in different subjects units: %.

Subjects \ Sources	Academic Journal	Master Dissertation	Doctoral Dissertation
Nature science	35.3	55.8	8.9
Social science	21.8	62.4	15.8

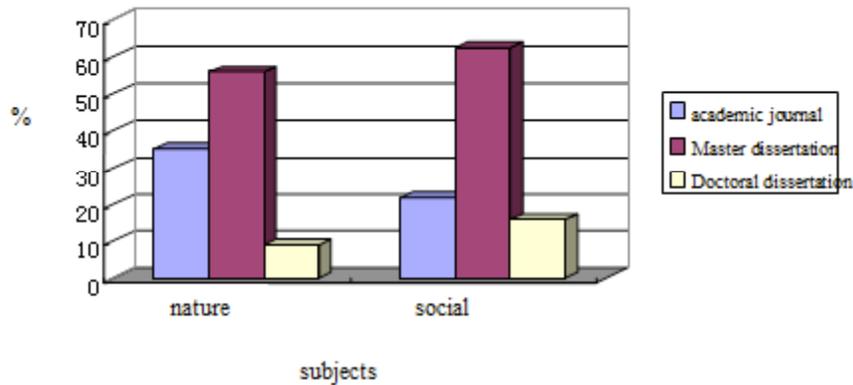


Fig. (5). Distribution state of the sources of the plagiarized literatures and the distribution of the number of plagiarism in different subjects.

Table 5 shows the sources of the plagiarized literatures and the distribution of the number of plagiarism among social and nature science papers.

Among the papers whose replication and reference rate over 20%, no matter the social science papers or nature science, all have largely plagiarize the papers of previous graduates. And their sources and plagiarism proportion roughly similar to each other, all concentrate upon master dissertation. For a wider perspective, nature science subjects plagiarize much more academic journal than social science subjects. Meanwhile, social science subjects obviously prefer

master and doctoral dissertation. It notes that different subjects rely on different literatures. As a result, we may take more concern about the literature categories and characteristics when educating the masters and focus on training their ability of literature research and analysis.

2.4. Impact of Tutors Time and Efforts on Replication and Reference Rate

In order to analyze the relationship between the two factors, we made a questionnaire in our college, handing out 300 questionnaires and acquire 300 effective questionnaires.

This research will analyze impact of tutors time and efforts on replication and reference rate.

To simplify the research, we supposed the tutors have made sufficient effort and time on master dissertation. Thus we select 137 papers whose replication and reference rate over 20% as samples and analyze the relationship between the quality of the papers and the tutors' instructions [6].

We assume that if a tutor is responsible for two or three students, then the effort and time invested on students would be half of and one-third of those who only instruct one student. This research conduct the correlation analysis of the data, here is the result:

Table 6. Detection results of the correlation of the initial replication and reference rate and time and efforts the tutors invested.

Variate	Initial Rate	Quantity of the papers
Initial rate	1	-0.324**
time and efforts invested	-0.324**	1

note: **means high level when P<0.01

The result shows that among the 137 papers, the replication and reference rate have a significant negative correlation with the tutors time and efforts. In other words, the more students the tutors have, the higher the replication and reference rate is.

According to the above analysis, we know that the replication and reference rate is relevant with the tutors time and efforts. The effort and time can be divided into five elements: first, the frequency of strengthening the academic norms; second, the length of the academic discussions; third, Guidance and help among the academic study; forth, frequency of correction; fifth, the frequency of the academic discussions. Hereon, through the statistical calculation of the data and modeling, we discussed the specific proportional relation between the replication and reference rate and the tutors time and efforts in details.

As there are many factors involved in the issue, we can obtain the relationship between these factors and the papers by utilizing the correlation, then solve and analyze the equation.

$$\epsilon_{0i} = \frac{1 + |s_0| + |s_i|}{1 + |s_0| + |s_i| + |s_i - s_0|} |s_0| = \left| \sum_{k=2}^{n-1} x_0^0(k) + \frac{1}{2} x_0^0(n) \right|, |s_i| = \left| \sum_{k=2}^{n-1} x_i^0(k) + \frac{1}{2} x_i^0(n) \right| \quad (1)$$

Table 7. Correlation degree of the different factors.

Replication and Reference Rate	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
grey correlation degree (r_i)	0.5735	0.7636	0.6414	0.6326	0.9465

$$|s_i - s_0| = \left| \sum_{k=2}^{n-1} (x_0^0(k) + \frac{1}{2} x_0^0(n)) + \sum_{k=2}^{n-1} (x_i^0(k) + \frac{1}{2} x_i^0(n)) \right| \quad (2)$$

Algorithm: Organize the related data and calculate the correlation to find the interplay between the current factors and charging standard of the college. We supposed when $r_i > 0.5$, we say the factor is related; when $r_i \leq 0.5$, we say the factor is unrelated. Here's the approach:

Utilize the grey system theory modeling software 3.0

Generation process of grey correlation

Star from calculating the absolute correlation of Sequence [1, 2].

Calculate $|s_0|, |s_i|, |s_i - s_0|$

First, orders the data into sequences

Sequence [1]: 4147.15, 4554.43, 4549.34, 6052.33, 6264.52, 6364.42, 6384.61

Sequence [2]: 14185, 16500, 20169, 23708, 25608, 30015, 35181,

Second, imaging the starting point of the sequence

Sequence [1]:

0.0000,407.2800,402.1900,1905.1800,2117.3700,2217.2700,2237.4600

Sequence [2]:

0.0000,2315.0000,5984.0000,9523.0000,11423.0000,15830.0000,20996.0000,

Third, calculate $|s_0|, |s_i|, |s_i - s_0|$

Conclusion: the correlation of Sequence [1] and Sequence [2] is 0.5735

-----the end-----

Calculated and finish the available Table 7.

Through the above analysis we get the result: $r_5 > r_2 > r_3 > r_4 > r_1 > 0.5$

We arrive at a conclusion that to ensure the quality of the master dissertation, the number of the students instructed by tutors should be determined according to the tutors time and efforts. Too many students may overload the tutors capability and leads to a high replication and reference rate as well as a lack of innovation.

But there are exceptions, in the research, two tutors have instructed five papers and replication and reference rate of eight papers below 20%, which means if a tutor is highly responsible and invest sufficient efforts and time, the master dissertation can achieve the standard even though the tutor is responsible for several students.

Table 8. Expectation and variance of thesis topic sources and the replication and reference rate.

	Optional Subjects	Subjects of science and Technology Plan Projects of Tangshan City	Projects Supported by Province	Projects Entrusted by Enterprises	Projects Supported by State	International Cooperation Projects
ratio	33%	11%	19%	8%	27%	1%
expectation	11.67	8.08	9.54	6.39	8.41	9.13
variance	106.14	84.87	49.98	25.89	59.94	66.90
quantity	73	24	42	18	58	3

Table 9. Difficulty coefficient and membership of the subjects.

	Optional Subjects	Subjects of Science and Technology Plan Projects of Tangshan City	Projects Supported by province	Projects Entrusted by Enterprises	Projects Supported by State	International Cooperation Projects
membership	0.013699	0.041667	0.02381	0.055556	0.017241	0.333333
difficulty	0.17	0.52	0.29	0.69	0.21	4.12

2.5. The Correlation of Topic Sources and the Replication and Reference rate

For topic selection: optional subjects 33%, subjects of Science and Technology Plan Projects of Tangshan city, Projects supported by the Natural Science Foundation of the Hebei province 19%, Projects supported by the Natural Science Foundation of the state 27%, projects entrusted by enterprises 8% and international cooperation projects 1%.

For the initial 1234 papers, there are 87 papers' replication and reference rate over 20%. Among which 73 papers are optional subjects, 2 papers are subjects of Science and Technology Plan Projects of Tangshan city, and 3 papers are Projects supported by the Natural Science Foundation of the Hebei province, 1 from projects entrusted by enterprises, 7 from Projects supported by the Natural Science Foundation of the state and 0 from international cooperation projects. The expectation and variance of thesis topic sources and the replication and reference rate, see Table 8.

The analysis of the topic sources aims at the master dissertation of 2014 spring graduates, and finds its characteristics according to the analysis of the level of the topic. In order to discover the dimensions of the plagiarism, we take the whole group as research object, thus to get rid of the negative effect of the fewer sample data.

2.5.1. Define the Items Level of the Thesis Topic Sources by Fuzzy Comprehensive Evaluation

The expectation and variance of thesis topic sources and the replication and reference rate, see Fig. (6).

Analyze the data of the Table 6, establishing the model of the difficulty subordinated function by utilizing fuzzy mathematics.

Here is the function:

$$f(n) = \frac{1}{d}, 0 < d < +\infty \tag{3}$$

To express the difficulty of the topic sources, we adopt the variation coefficient method to handle the data and select the ratio of the coefficient and averages.

$$H = \frac{f(n)}{E} \tag{4}$$

Apply the data into the function and get the difficulty coefficient and membership, see Table 9.

2.5.2. Discuss the Correlation Between the Expectation and Variance of Thesis Topic Sources and the Replication and Reference Rate on the Basis of Grey Relational Analysis

To discuss the correlation of the replication and reference rate thesis topic sources, we can calculate the data through relevancy. Establish and solve the grey correlation.

Organize the related data and calculate the correlation to find the interplay between the difficulty of the topic sources and the expectation and variance. We supposed when $r_i > 0.5$, we say the factor is related; when $r_i \leq 0.5$, we say the factor is unrelated, the data is unavailable.

After calculating, we get the correlation of the difficulty of the topic and the expectation of the replication and reference rate $r_1 = 0.8465$, the variance correlation $r_2 = 0.6414$

That is to say $r_2 > r_1 > 0.5$

Conclusion:

Through the analysis we find the expectation goes down with the growth of the subjects' difficulty which means the replication and reference rate is lower. The expectation of optional subjects is 11.67, the highest among all the subjects; the lowest expectation is 6.39, they are projects entrusted by enterprises.

For variance, from the highest 106.14 of optional subjects and the lowest 25.89 of projects entrusted by enter-

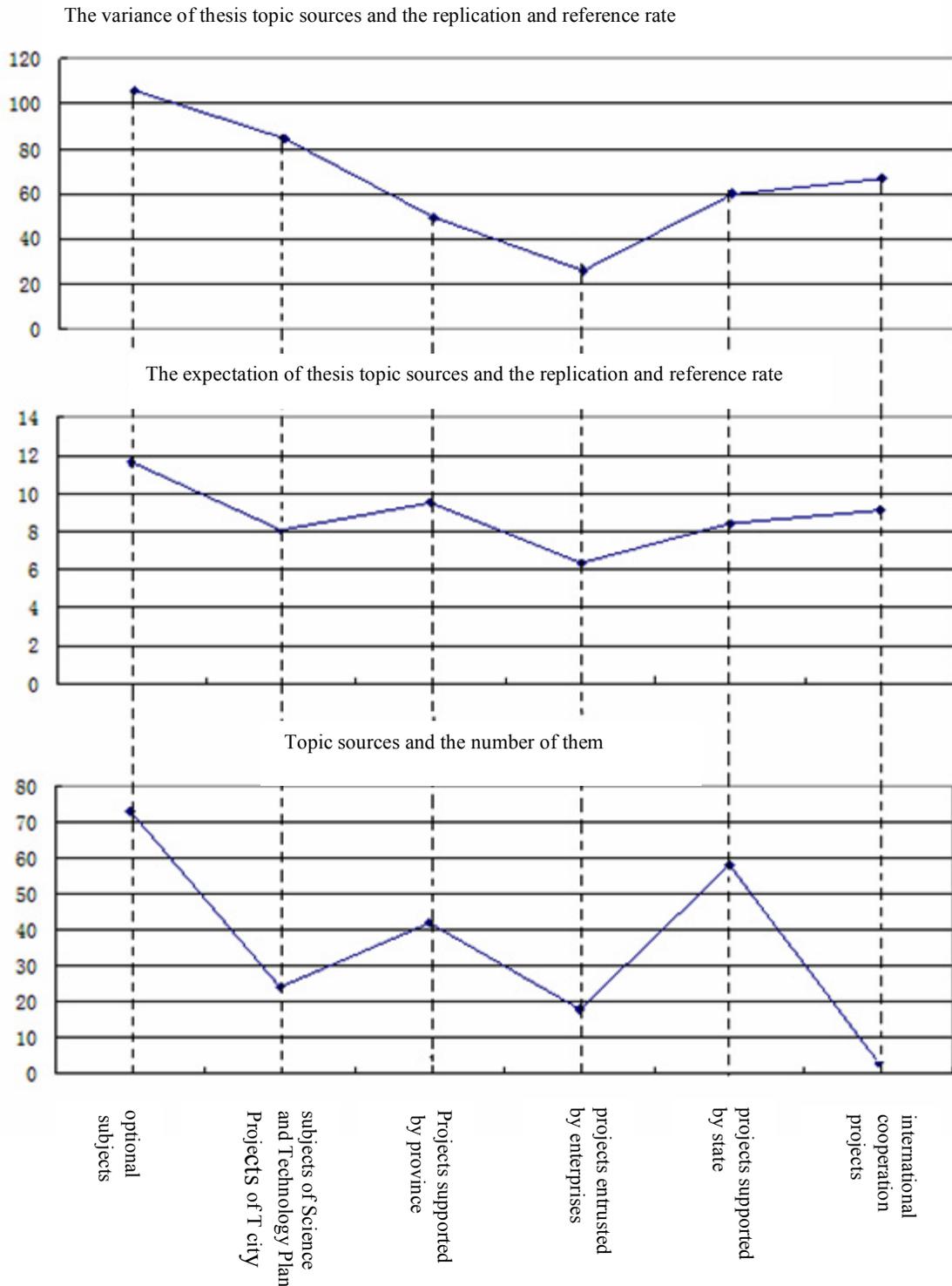


Fig. (6). Expectation and variance of thesis topic sources and the replication and reference rate.

prises indicates the distinction of masters’ academic research capability.

Some graduates own an attitude of following herd that has chosen the optional subjects, and this is the current status of the master dissertation [7]. On contrast, projects entrusted by enterprises request practical demand and requirements so it greatly propel the motivation of innovating in papers writing.

General performance:

1. The higher subjects’ level is, the lower replication and reference rate is. The optional expectation is the highest and the projects entrusted by enterprises are the lowest.
2. The higher subjects’ level is, the higher the graduates’ academic capability is.

Table 10. Replication and reference rate distribution in each chapter of the master dissertation units: %.

Chapter	Text	Literature Review	Third Chapter	Forth Chapter	Others
Number proportion	86.5	93.8	62.5	37.9	27.0

2.6. The High Sectors of the Replication and Reference Rate of Medical Thesis Concentrate upon the Section Literature Summarization and the Section Text.

This research focus on high replication and reference rate, namely the 45 medical papers with a replication and reference rate over 30% and separating and determining the section Literature Summarization and the section Text. After comprehensive analysis, we draw a conclusion that the Literature Summarization section is the high risk area of plagiarism, see Table 10.

Fig. (7) shows the serious plagiarism happened in the 45 papers and the rate amounts to 86.5%, each paper’s plagiarism above 7000 words, some even over ten thousand words. Besides, more than 5000 words plagiarize in Literature Summarization section of 38 papers. The plagiarize phenomenon reduced after chapter 3. So we know the plagiarize issues are pretty serious in the section Literature Summarization and the section Text.

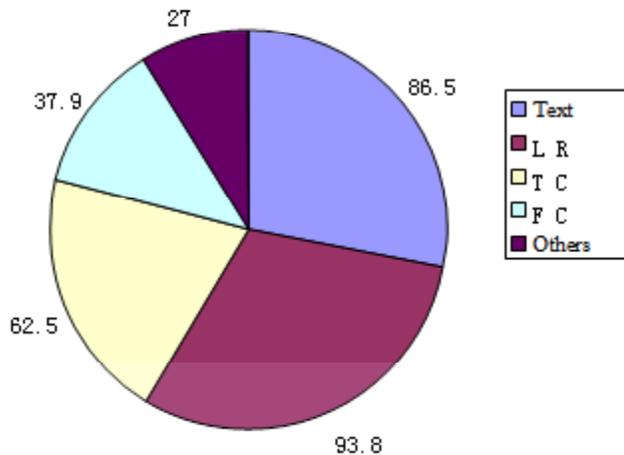


Fig. (7). Replication and reference rate distribution in each chapter of the master dissertation.

We investigate why this occurred [8]: some masters may do not know the norms of the quotation and note the quoted views, concepts or results; or they just cobbled together the essay without indicating the sources. All these factors lead to

the high replication and reference rate of section Literature Summarization and the section Text.

CONCLUSION

This paper focus on 1234 master dissertation applying for master degrees in HUU in 2014 and do the statistical analysis of the initial replication and reference rate detection conclusion. Thus to illustrate the plagiarize phenomenon and its features and provide suggestions for restrain academic misconducts. Furthermore, we are supposed to combine the results and relevant analysis with our college’s master education so as to better reflect and extend the function of our degree management department and set sound standards of academic moral conducts for our star-up training site for master's and doctor's degree.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] E. Wang, “Academic norms education in university,” *Journal of Changchun University of Technology*, vol. 31, no. 1, pp. 2-3, 2010.
- [2] G. Ma, “The performance, causes and countermeasures of the anomie of academic morality among postgraduates,” *Journal of Shanxi Finance and Economics University*, vol. 31, no. 2, pp. 296-297, 2009.
- [3] Z. Xiong, and R. Fang, “A model study on the main influence factors on replication and reference rate of master dissertation,” *Journal of Graduate Education*, no. 4, pp. 40-45, 2011.
- [4] R. Fang, P. Guo, and D. Fang, “A Model Study on the Self-influence Factor on Academic Behaviors Among Graduates,” *Journal of Higher Education Management*, vol. 6, no. 4, pp. 84-91, 2012.
- [5] R. Fang, D. Fang, and P. Guo, “Features of academic misconducts on master dissertation,” *Academic Degrees & Graduate Education*, no. 5, pp. 18-22, 2013.
- [6] B. Jin, “Plagiarize the work of others is a serious infringement—the cognizance of plagiarism,” *Shan Xi Library Journal*, vol. 4, 2001.
- [7] X. Zhang, and F. Ming, “A Study on the Relationship Between Academic Norms and Intellectual Property Rights in Graduate Education,” *Science and Technology Management Research*, vol. 12, 2010.