

PLAGIARISM ACROSS EUROPE AND BEYOND

Conference Proceedings

**June 12–13, 2013
Brno, Czech Republic**

Advisers and reviewers

Professor Tracey Bretag
Dr. Erik Borg
Ms. Jude Carroll
Dr. Elizabeth Cox
Professor John Davies
Professor Catherine Demoliou
Dr. Lisa Ganobcsik-Williams
Dr. Melpo Iacovidou
Professor Krzysztof Jóźwik
Dr. Angelika Kokkinaki
Mrs. Agnieszka Michałowska-Dutkiewicz
Ms. Gill Rowell
Dr. Linas Stabingis
Dr. Debora Weber-Wullf

Organization committee

Tomáš Foltýnek—head of the organization committee
Dita Dlabolová—communication
Irene Glendinning—review process coordination
Petra Pečínková—supporting programme
Jiří Rybička—proceedings typesetting
Pavel Turčínek—technical support

Organizing committee is most grateful to the IPPHEAE project team members for their support and advice. Special thanks to all reviewers for their hard work.

Table of contents

Irene Glendinning: Preface—About Project	5
Keynote Presentation Abstracts	7
Tracey Bretag: Exemplary Academic Integrity Project: Key lessons for Australia, Europe and Beyond	8
Jude Carroll: What we can and cannot learn from each other about managing plagiarism	9
Jiří Janoušek: Anton summary	10
Gill Rowell: Addressing student plagiarism in the UK, ten years on	11
Debora Weber-Wulff: Plagiarism in German Doctoral Dissertations: Before and beyond zu Guttenberg	12
Papers—Section I	13
Blanka Farková: Intertextuality in Student Works—Comparison of Results of the Swedish and the Czech Study	14
Stephen Gow: A cultural bridge for the academic concept of plagiarism: A comparison of Chinese and British cultural concepts of plagiarism by Chinese Master's graduates of UK institutions employed by Sino-foreign joint ventures in Shanghai, China	27
Erja Moore: Sloppy referencing and plagiarism in students' theses	42
Stella-Maris Orim, Erik Borg, Isabella Awala-Ale: Students' experience of institutional interventions on plagiarism: Nigerian case	54
Neil Wellman, Julia Fallon: International MBA Students' academic malpractice: a quantitative survey	70
Papers—Section II	92
Muna Alsallal, Rahat Iqbal: An Approach to Detect Illegal Similarity in Research Literature Using Latent Semantic Indexing	93
Robert Craig, David Dalton: Understanding first year undergraduate student perceptions of copying and plagiarism: developing a platform for a culture of honest inquiry and the academic construction of knowledge	103
Sharon Flynn: Teaching staff concerns about academic integrity and their implications for staff development	115
Tomáš Foltýnek, Jiří Rybička, Catherine Demoliou: Do students think what teachers think about plagiarism?	127
Adrian Lee, Antony Edwards: Assessing the value of a holistic use of Turnitin to promote academic integrity	136
Shiva D. Sivasubramaniam: Electronic plagiarism detection software as self-teaching tool for plagiarism avoidance in Bioscience undergraduates	149
Rui Sousa-Silva: Investigating Academic Plagiarism: A Forensic Linguistics Approach to Plagiarism Detection	159
Papers—Section III	170

Eckhard Burkatzki, Joost Platje, Wolfgang Gerstlberger: Cultural differences regarding expected utilities and costs of plagiarism between high trust- and low-trust-societies—preliminary results of an international survey study	171
Angelika Kokkinaki, Melpo Iacovidou, Catherine Demoliou: Students' Perceptions on Plagiarism and Relevant Policies in Cyprus	192
Július Kravjar, Juraj Noge: Strategies and Responses to Plagiarism in Slovakia	201
Ľuboš Lunter, Daniel Jakubík, Šimon Suchomel, Michal Brandejs: Inter-university Cooperation on Plagiarism Detection Systems in Czech Republic	216
Anna Michalska: Students and staff voices on “zu Guttenberg’s case” and its influence on plagiarism awareness in German HEIs	225
Aurelija Novelskaitė, Raminta Pučėtaitė: Plagiarism in Lithuanian academia: formal definition and informal attitude	236
Papers—Section IV	248
Daniela Chudá, Jozef Lačný, Maroš Maršalek, Pavel Michalko, Ján Súkeník: Plagiarism detection in Slovak texts on the Web	249
Tomáš Kučečka, Daniela Chudá: Improving Plagiarism Detection through the Document Content Segmentation	261
Ondřej Veselý, Jan Kolomazník: Predicting number of search engine results to optimise online plagiarism detection	270
Workshops, Posters	278
Irene Glendinning: Yes, they are all plagiarising, but what can be done to stop it?	279
Marcelo Krokosz, Reinaldo Putvinskis: Analysis of the perceptions of undergraduate students in Business Administration on the occurrence of academic plagiarism in Brazil	281
Trudy Somers: Academic dishonesty and extenuating circumstances in the global classroom	283
Tracey Bretag, Gill Rowell and Teddi Fishman: Academic integrity in higher education: A global perspective	287
Peter Kročitý: Applying principles of scholastic honesty at Vysoká škola manažmentu v Trenčíne	288
G. Matheou, A. I. Kokkinaki and C. Demoliou: On Anti-Plagiarism in an Open University: Challenges and Best Practices	289

PREFACE—ABOUT PROJECT

Irene Glendinning

Impact of Policies for Plagiarism in Higher Education Across Europe

The project Impact of Policies for Plagiarism in Higher Education Across Europe (IPPHEAE) is the motivation behind this conference. The project is funded by the European Union's Lifelong Learning Programme and runs between 1st October 2010 and 30th September 2013.

Project Number: 510321-LLP-1-2010-1-UK-ERASMUS-EMHE

Sub-programme: Modernisation of Higher Education

The project is focused on strategic responses to plagiarism and academic dishonesty, nationally and institutionally, as evidenced in Higher Education Institutions across the 27 countries of the European Union, focusing specifically on policies and procedures applied at bachelor and masters levels. The team is interested in exploring examples of good practice in both detection and prevention of plagiarism.

The aims of the project are to

- Discourage student plagiarism by developing knowledge and skills in all participants;
- Promote greater understanding of author's rights, plagiarism prevention policies and academic writing practices;
- Explore and compare current expectations of teachers and students for standards in academic writing;
- Compare how HEI policy and procedures for handling plagiarism are established and how they are operating in HEIs across the EU;
- Encourage use of readily available resources for prevention of plagiarism;
- Involve as many participants as possible in the research and dissemination activities;
- Improve standards and quality in higher education across the EU and beyond.

The objectives of the project are to

- Explore, develop and evaluate information resources, tools and methods for plagiarism detection and penalties;
- Conduct an EU-wide survey to explore national and institutional responses to plagiarism and academic dishonesty;
- Document cases of good practice in plagiarism prevention;

- Evaluate effectiveness and impact of policies for detection and prevention of plagiarism
- Exploit the findings of the research to disseminate good practice;
- Make recommendations for improving policies and procedures for responding to the threats of plagiarism and academic dishonesty in higher education across the EU and beyond.

Lead Partner:

Coventry University, United Kingdom, Principal Investigator: Irene Glendinning
email: ireneg@coventry.ac.uk

Project Partners:

- **Aleksandras Stulginskis University**, Lithuania, Coordinator: Dr Linas Stabingis, email: linas.stabingis@asu.lt
- **Mendel University**, Czech Republic, Coordinator: Dr Tomáš Foltýnek, email: foltýnek@pef.mendelu.cz
- **Technical University of Lodz**, Poland, Coordinator: Agnieszka Michałowska-Dutkiewicz, email: agnieszka.michalowska-dutkiewicz@p.lodz.pl
- **University of Nicosia**, Cyprus, Coordinator: Dr Catherine Demoliou, email: demoliou.c@unic.ac.cy

Project Consultant:

Jude Carroll

Project Sponsors:

Turnitin/iParadigms (www.turnitin.com) and IS4U (www.is4u.cz)

Project web site:

<http://ippheae.eu>

KEYNOTE PRESENTATION ABSTRACTS

EXEMPLARY ACADEMIC INTEGRITY PROJECT: KEY LESSONS FOR AUSTRALIA, EUROPE AND BEYOND

Tracey Bretag

Abstract: The call to examine consistency in academic integrity in the United Kingdom came from the Independent Adjudicator for Higher Education and resulted in the development of the project, Academic Misconduct Benchmarking Research (AMBeR). The AMBeR Project looked at the range and spread of penalties available for student plagiarism among UK higher education institutions. The Impact of policies for plagiarism in Higher Education Across Europe (IPPHEAE) project was funded by the European Union and has investigated the policies and procedures for detecting and preventing student plagiarism in Higher Education Institutions (HEIs) across Europe. In Australia, the Tertiary Education Quality and Standards Agency (TEQSA), following the work of the Australian Universities Quality Agency (AUQA), made clear its expectation that Australian higher education providers should have “systematic, mature internal processes for quality assurance and the maintenance of academic standards and academic integrity”. It could be argued that by failing to adequately confront/prevent lapses in academic integrity, many HEIs across the globe are neglecting a broader duty to society to ensure that students are learning rather than cheating and that they have in fact earned their degrees. When academic integrity breaches go unchecked, this has the potential to undermine the credibility of degrees and the reputations of educational providers. Furthermore, society as a whole suffers because it is difficult for employers to determine who is and is not qualified and this potentially puts at risk the people who rely on well-trained professionals such as doctors, nurses, lawyers, pilots, engineers and teachers, to mention a few. While individual HEIs are interested in protecting their own reputations, it is imperative for the credibility of the higher education sector as a whole that they are seen to embrace a proactive and reliable approach to academic integrity, particularly in light of international student mobility and internationalisation more generally.

WHAT WE CAN AND CANNOT LEARN FROM EACH OTHER ABOUT MANAGING PLAGIARISM

Jude Carroll

Abstract: I have spent more than a decade in working and travelling around Europe to many different countries, to deal with student plagiarism. Some countries manage plagiarism well but many put little or no effort into dealing with plagiarism. I will describe some approaches I have seen or suggested which seem to be useful across many contexts. And I will describe some which go nowhere, leading to local frustration and student confusion. Often, dealing with plagiarism gets “stuck” because of local issues, beliefs and regulations. The purpose of the keynote is to suggest how, when and whether, in the future, people can address the issue of student plagiarism, based on others’ experiences, data and stories.

ANTON SUMMARY

Jiří Janoušek

Abstract: The presentation is aimed at AntOn solution which was created as a key output of IPPHAEA project. The AntOn stands for “Antiplagiator Online” and the solution was developed by IS4U company. AntOn is extending the portfolio of tools and solutions offered by IS4U company’s main product UIS—the complex information system. The UIS system provides administrative support to schools and there assists in many key activities—e.g. administrative, study and research processes support. One of the key activities is the research and working with student texts, and that is why the company has decided to create AntOn when asked to participate at IPPHAEA project. The AntOn solution was successfully developed and launched during the project and it was tested together with UIS system in production.

The AntOn solution is a special software component executed at a standalone server. It is intended to accept any texts (documents) in various formats (e.g. DOC, DOCX, PDF, HTML, XML, ODT) to analyze them and to report back about texts’ level of similarity to the other texts in the AntOn internal document database. The REST interface is implemented to connect AntOn to any system as a service—e.g. REST connection is used for cooperation with UIS, too.

The AntOn itself works in 4 steps. Firstly, the plain text is extracted from the input document. Secondly, language-dependent stop lists and special language settings are applied. Stop list is a group of extra words, fillings or frequently used words, numbers, special characters etc. These words have no importance for the text processing and they can be removed. Special language settings ensures the most appropriate output for certain language. Next step includes creating special five-words pieces called chunks. A chunk is context-dependent surroundings of the word, so it needs to be created for almost every word (exactly it is $n - 4$ chunks for n words) in the processed text. Finally, the set of chunks for processed document is compared to other chunk sets in the AntOn document database. This is computationally very demanding operation, so there is a special approximation algorithm implemented. It is vital to have good hardware capacity to achieve acceptable speed and results.

The output of the AntOn system is level of similarity of processed document to other documents in AntOn’s database. High precisiuous SDIFF algorithm can be used to show differences between two similar documents when the level goes over the given limit (now the limit is set to 50%). The final decision about plagiarism, however, needs to be made by human inspector, the AntOn system provides only the level of similarity.

To conclude, it is necessary for the school to have any antiplagiator solution because just the existence of the antiplagiator prevents students from cheating. When some survey in existing antiplagiators has been done it was found out that the quality of solution is not crucial. The school is satisfied and is not forced to increase the accuracy of the system. Therefore, the IS4U Company implemented AntOn system at full stable version and due to this no further development is considered now.

More information can be given at info@is4u.cz.

Author

Jiří Janoušek, jiri.janousek@is4u.cz, IS4U, s. r. o., U vodárny 2a, Brno, Czech Republic

ADDRESSING STUDENT PLAGIARISM IN THE UK, TEN YEARS ON

Gill Rowell

Abstract: UK higher education institutions have been on a 10 year long journey to establish a clear and effective response to student plagiarism and develop a proactive model of academic integrity awareness. Created as a government-funded project in 2002 against a background of growing concerns over the originality of student work the Plagiarism Advisory Service (forerunner of PlagiarismAdvice.org) sought to ignite institutional thinking on plagiarism, underpinned by considered and integrated use of the Turnitin plagiarism detection software.

During this time there have been many challenges and triumphs, with many institutions developing academic integrity policies and procedures where none previously existed and rethinking the whole assessment and student learning process.

Accordingly general awareness of the problem in the UK, although difficult to accurately gauge, has improved dramatically and policies and guidelines are in place at most universities. Turnitin is employed by 98% of UK universities and there is recent evidence to suggest that instances of unoriginal content as flagged by Turnitin are decreasing.

Alongside these successes the last decade has also confirmed what we have always known about plagiarism and how we address it, ie it's complex; no one size fits all solution exists and most definitely, technology alone isn't the answer.

Author

Gill Rowell, Academic Advisor, Turnitin/PlagiarismAdvice.org

PLAGIARISM IN GERMAN DOCTORAL DISSERTATIONS: BEFORE AND BEYOND ZU GUTTENBERG

Debora Weber-Wulff

Abstract: In 2011 activists at the GuttenPlag Wiki publically documented the extensive plagiarism in the dissertation of the German Minister of Defence, Karl-Theodor zu Guttenberg. The final tally—63% of the lines on 94% of the pages—was shocking and he was forced to resign, but this was neither the first nor the last doctoral dissertation in Germany to be found to contain plagiarism.

In 1865 the Faculty of Philosophy of the University of Marburg published a broadsheet with the names of two persons caught plagiarizing in their dissertations. The Faculty of Philosophy of the University of Bonn decided in 1965 to rescind the doctorate in German Literature of a Dominican Order priest. The college of deans voted, however, not to depromote him for reasons still unknown.

In 1990 another case at the University of Bonn was widely published. One professor published an extensive documentation of the plagiarism in another s dissertation. In 2005 there was a case at the University of Tübingen, at the Faculty of Catholic Theology, that was widely reported on. Many other cases, however, have not been discussed publically at all.

Since zu Guttenberg there have been many documented cases of plagiarism. VroniPlag Wiki has documented over 30 cases of plagiarism. Some universities have rescinded the doctorates on the basis of the documentations, but others have refused to even publish explanations as to why the extensive text parallels are not considered problematic.

This talk will present both historic and current cases.

PAPERS—SECTION I
INTERNATIONALISATION, STUDENT MOBILITY AND
ACADEMIC INTEGRITY

INTERTEXTUALITY IN STUDENT WORKS—COMPARISON OF RESULTS OF THE SWEDISH AND THE CZECH STUDY

Blanka Farková

Abstract: During the period from September 2012 to November 2012, the research focused on the evaluation of intertextuality in student works by the academics took place in the Czech Republic. Methods objectives and the research plan from the Swedish research study by Pecorari and Shaw (2010, 2012) were adopted for the purpose of implementation of the research. Overall there were implemented eight semistructured interviews proceeded on five textual extracts with respondents from two universities. Complete results of the Czech study will be published in another contribution.

This contribution presents the results of the comparison of both independent studies (the Czech one and the Swedish one). The aim of the comparison is to find out and analyze compliances and differences in the evaluation of the intertextual links between Czech and Swedish academics. The results have shown that Czech academics are more conflicting in their opinions on intertextuality than Swedish respondents. Czech respondents labeled presented extracts as plagiarism more often than Swedish academics.

Key words: plagiarism, intertextuality, source use, academic writing, Czech Republic, Sweden

Introduction

In the Czech Republic there has not been any research conducted targeted at findings, documentation and analysing of reactions of academic staff on various intertextual links between a student work and the source. However some research of this type was carried out in Sweden. At the same time, Swedish researchers Pecorari and Shaw (2010, 2012) tried to record interpretations that these academics tied to student texts. They wanted to document differences in their approaches to the assessment of intertextuality. In the article “Types of student intertextuality and faculty attitudes” Pecorari and Shaw (2012) published methods they used for the research of the evaluation of intertextual relations in student works by the academics. With agreement from the authors we adopted these methods and used them in the scope of our research.

Because we used identical methods, we could compare our findings with published results of the Swedish study (Pecorari, Shaw 2010, 2012). The aim of the comparison is to find out and analyze differences in the evaluation of intertextuality by Czech and Swedish academics.

We expected that the diversity of opinions and evaluation of respondents would be captured. Our expectations were based in particular from foreign research and professional sources which do not focus just on the diversity of opinions of academic staff on defining plagiarism (closer Flint, Clegg and MacDonald 2006; De Jager, Brown 2010; East 2010; Roig 2001; Borg 2009).

The contribution compares the views of Czech and Swedish academics on identical textual extracts. Examination of consensus or diversity is beneficial mainly in case

of comparative examination of plagiarism. Findings can be effectively taken into consideration in mutual cooperation and practical implementation of projects between both countries.

The Method

Methods of the Swedish research were transferable into Czech environment without any necessary editing. In order to protect the copyright, however, it was not possible to use the fifth textual extract for the implementation of the study in Czech environment. The following part of the text describes not only the original Swedish method, but also the Czech modification.

The Swedish research was carried out with the method of semistructured interviews in English language. In contrast to the Swedish research, all interviews were conducted in Czech language and presented texts were also translated into Czech. None of the respondents opted for the assessment in English language. This difference may be because the Swedish research was conducted by six Swedes and two academics who moved to Sweden from another European country (interviews were read in Swedish \times English in proportion 5×3). In our case, all respondents were Czech. There were eight interviews in total evaluated in our study. It was the same number of respondents similarly as in case of the Swedish research.

Interviews were recorded with the consent of the respondents. The length of interviews was in the range from 15 minutes to 30 minutes. Here you can find a difference between the Czech and the Swedish research where the length of the interviews was between 25 and 60 minutes. The diversity was probably caused by the fact that Czech academic staff did not actively come up with new themes during the interview. Their reactions were always directed to the asked question. The interview scenario was used according to the Swedish research, which is attached in the Appendix. The extracts were presented to respondents gradually, from the extract number one up to the extract number five. This was followed by a part devoted to all extracts at the same time (their comparison and so on). In the last part of the interview respondents were asked more general questions whose objective was to find out the opinion of participants of the related area from more general point of view.

In case of each extract, first of all a question, was asked that required approving or disapproving answer. The question was directed to the assessment of acceptability or non-acceptability of the way of use of the source and of referring to it. This was followed by questions focused on the justification of the assessment and following steps which were raised by this assessment (actions against the student, explanation of the problem to the student, and so on). If the respondent himself did not comment whether it was a case of plagiarism, he was asked about this fact. The question itself was again built the way the respondent would take a firm and clear viewpoint. Then questions were asked that related to the justification of his or her opinion.

Within the work with all extracts, the respondent was asked to compile the examples into continuum from completely acceptable to completely unacceptable way of application and referring to the source. Questions about explaining the placement of individual examples into the field of the continuum followed.

Table 1
Characteristics of Examples

Example	Number of words	Matching words	The longest chain of matching words	Signaling of citation	Reference to the source	Nature of the content
1	37	12	4	No	No	General
2	28	22	11	No	Yes	Results
3	36	29	29	No	Yes	General
4	27	26	13	No	No	General
5 (original)	311	285	121	No	Yes	Results
5 (amended)	80	77	44	Yes*	Yes**	Definition

* citation of the secondary resources,

** Reference to the secondary source

The Structure of Assessed Texts

Pecorari and Shaw used for their research such examples which included various intertextual relations. Examples differentiated among themselves as follows (overview of characteristics of examples is in Table 1):

- length (extracts, different length of transmitted words and concatenation of words),
- referring (signalized quotation, reference to the source),
- signaling of secondary citation (referring to the source to which the source material of the student refers),
- type of reproduced text (transmission of general information, transfer of research findings and definitions).

So the participants could recognize relations in the word transmission (individual and concatenated) more easily, the same or similar words and concatenated words in both researches were marked in red. Conformity of transferred words moved approximately from less than one-third of words from the source text (example number one) up to much higher transmission, or rather up to full transmission (example number five). Different rates of adaptation of the information transmission also appeared in the examples. Examples number one and four do not contain the reference to the source. Other examples include the reference, however, in one case it is only references to secondary resources. Some examples transmit the original thoughts, while some of them transmit generally well-known information. The last fifth extract of the text has not been adopted with respect to the Swedish copyright. We used an extract of a text by a Czech student and traced necessary original sources and the source text. This sample has the same characteristics which were assessed in the Swedish study.

The examples were as from the field of natural sciences (biology—examples number 1–4), but also from the field of humanities (special pedagogy—example number 5).

Some respondents were closer to the content of the sources which students used for the creation of their works, and the others were clearly far beyond the topics which were captured in works of students. All of them were, however, acquainted with generally applicable standards for the recognition of authorship.

Respondents

Our respondents, as well as Swedish respondents, differed in their academic position, as well as in the length of academic practice. We deliberately left this criterion up to the selection of respondents, because the Czech academic staff, in scope of all academic positions and with different length of practice, can lead qualification works of students. They can also be members of the examination committee during the state exams, or they can also enter the disciplinary proceedings brought against the student suspected of plagiarism and so on.

Swedish respondents were settled in the field of natural sciences, engineering or medicine. Our respondents worked both in natural sciences and in humanities. We enriched the selection with the respondents from the field of humanities. We wanted to find out what kind of progress will appear in the evaluation made by academic staff from a completely different scientific discipline, and what factors of evaluation will enter their assessment.

Respondents were anonymous in relation to the researcher, as the recording of the interview was conducted by a trained interviewer who was choosing and asking the academics according to the key of the selection of respondents. The only selection criterion was the voluntariness of the academic participating in the research. The researcher was provided only with the tape with recorded interview which was marked only with the serial number of the respondent. Swedish researches took place anonymously in relation to the public. Respondents and researchers knew each other personally, but the identity of participants was anonymous in relation to the professional public—which means during the publishing of results of the study (see Table 2 with the attached list of participants in the Czech and the Swedish research).

The interviews were recorded, as has already been mentioned, through the independent interviewer. This way of recording the interviews was carried out because the author of this work herself was a student. If the recording of the interviews had been carried out by the author herself, then the research could have been depreciated with several possible handicaps. The academic staff may not have been willing to participate in the research (research was carried out on a voluntary basis). Also the respondents may not have been helpful in their replies. For this reason questioners were selected who had the status of an academic. The questioner had experience with leading the thesis. He also dealt with cases of plagiarism in qualification works of students in the past. This method of implementation of interviews was also selected because if the author of work had recorded the interviews herself and had subsequently evaluated them, there might have appeared a subjective distortion of the obtained data. On the contrary, interviews in the Swedish research were realized by the authors themselves on respondents they personally knew. Therefore, it can be summarized that, in our

Table 2

Participants on the Research¹

Respondent		Academic position		Sex		Discipline	
Victoria	1	PhD student/tutor	Assistant professor	Female	Male	Environmental science	Politics
Kjell	2	Reader	Research worker	Male	Male	Zoology	Social and mass communication, microbiology
Gregorija	3	Reader	The Head of Department	Female	Male	Computer science	Social and mass communication
Karin	4	Senior lecturer	Associate Professor	Female	Male	Health care	Sociology and history
Lars	5	Full professor	Doctor	Male	Male	Biology	Social and mass communication, the environment
Mikael	6	Professor emeritus	Associate Professor	Male	Female	Biology	Pedagogy
Urban	7	Full Professor	Research worker	Male	Female	Biology	Pedagogy
Stefan	8	Full Professor	The Head of Department	Male	Male	Biology	Philosophy

¹ For the part concerning the Swedish research, the resources for the creation of the table were taken from the Table 1, Pecorari, Shaw (2012, p. 151).

Explanatory Note:

The Swedish Research 2010

The Czech Research 2012

Note: Associate Professor is an academical-pedagogical rank of academia granted in the Czech Republic

research, we managed to exclude other factors, which could cause devaluation of the results.

Process

During September and November 2012, semi-structured interviews with eight academics, who also work at the Charles University in Prague and at the University of Economics in Prague, were carried out. The Swedish research, by contrast, was conducted in spring 2010. Time delay between particular researches was not too significant. It is necessary, however, to consider whether the new findings in the field of plagiarism in this period could influence the evaluation of Czech academic staff. Both Czech educational institutions had not been significantly involved in the projects and conference activities on this theme. Therefore, we can assume that interviewed academic staff did not need to be informed about the results of the Swedish research.

The Results and the Discussion

Two main themes which appeared in the Swedish research were the following:

- participants had different and conflicting views of the acceptable and the unacceptable intertextuality,
- and they differed in factors which they considered while answering the questions.

Both of these themes were also reflected in the Czech research, even if, in some cases, in different nuances.

Questions were built the way that respondents should essentially agree or disagree. As in the Swedish research, Czech academic staff also answered hesitantly and evasively in many cases. Some of our respondents even answered to some questions by presenting general views of the discussed area, and they identified themselves with these views. The Swedish as well as the Czech respondents both reported out loud what they were thinking about. However, dissimilarity was recorded in case of some academics who answered questions without hesitation and directly with a clear answer, which was not detected in case of Swedish respondents.

Conflicting Views

The answers of Swedish respondents were often tentative and some of them mutually opposed. In case of Czech respondents, the conflict appeared in a different form. Some respondents presented their assessments. In the course of time, various factors were entering these assessments. Under the influence of a new factor which they had taken into account, they subsequently modified or completely changed their assessments. Therefore, the evaluation of certain specific respondents did not just move within the less or more acceptable level, but it also went across the continuum of acceptability—unacceptability in both directions. It was similar in case of statements considering the plagiarism (whether or not it was a case of plagiarism in a specific case).

In the Swedish research all of the respondents evaluated the fifth example as non-acceptable. No other example got such an identical evaluation. By contrast, in the Czech research six respondents identified the fifth case supplemented by ourselves as completely unacceptable, and two respondents as an example totally acceptable. In comparison, this non-compliance might have been caused by the fact that we did not select an applicable example. We are aware of that fact, but it should be noted that the results for other examples also vary significantly, and not only in comparison with the Swedish research, but also in the case of Czech respondents themselves. Consequently there is nothing signaling that example number five was selected inappropriately. We can demonstrate all this on the example number three. This example was found unacceptable by Swedish respondents in five cases. In case of Czech participants the results varied from two entirely acceptable evaluations to the unacceptable evaluation in case of one respondent.

In case of the Swedish research, the acceptability appeared most often in examples one and two and unacceptability mostly in case of examples three and four. In case of Czech respondents it can be only stated that example number four was moving on or beyond the borderline of acceptability. The following Table 3 shows

Table 3

Evaluation of Acceptability of Examples by Swedish Respondents¹

Respondent	Acceptable	“Border” or Uncertain Statement	Not acceptable
Victoria	2		1*, 3*, 4*, 5*
Kjell	1, 2		3*, 4*, 5*
Gregorija		1, 2	3, 4, 5
Karin	1		2, 3, 4, 5*
Lars	1, 2, 3	4	5
Mikael	1, 2	3	4, 5
Urban	1, 2	3	4, 5
Stefan	1, 2		3*?, 4*?, 5

¹ Modification of Table 3 taken from Pecorari and Shaw (2012, p. 153).

The explanatory notes: * examples which were identified as plagiarism by respondents;

? uncertain statement, border case

Table 4

The Continuum of Acceptability in Case of Czech Respondents¹

Respondent	Acceptable—Unacceptable				
1.	1	2*	3	4?	5
2.	3	1*	2	4	5*
3.	3	2	1	4	5
4.	2	3	1	4	5
5.	2	1	3	4	5
6.	1	3	4	2	5
7.	5	1	3	2	4
8.	5	2	1	3	4
The explanatory notes:		Plagiarism			

¹ Table shows the final evaluation of respondents.

* no assessment concerning the fact whether or not it was a case of plagiarism;

? uncertain statement, border case

examples acceptable and unacceptable as they were reported by Swedish respondents. Furthermore, there is the Table 4 which reflects the continuum of examples from fully acceptable to fully unacceptable as the respondents themselves stated during the interview. The data on the continuum of the Swedish research are not available.

As it is evident from the tables above, Czech academics are more conflicting in their opinions of what constitutes acceptable and unacceptable intertextuality than Swedish respondents. Czech academics also more often indicate unacceptable cases as plagiarism.

The Decisive Factors

As the decisive factors we indicate facts on basis of which the respondents made the decision about acceptability and unacceptability of the example. In this part of the text we describe how these factors entered the assessment of respondents. Particular examples of the factors are then stated in additional parts of the text (legitimate and illegitimate intertextuality).

Swedish respondents discussed their considerations whether it was an acceptable or unacceptable case, and they reported the factors which they took into account during their assessment. Swedish researchers came up with an interesting finding that individuals responded differently to the same specific case. If they responded differently, they logically explained the reason, why they had changed their mind. They also differed in their opinion of which intertextual features granted legitimacy to the text.

Similarly as in the case of Swedish research, our respondents also reported factors which they focused on during their assessment. The following factors were emerging depending on the problems which were identified in the submitted examples by the respondents. If the respondents had not identified any problem in a student text, they only stated that the example was acceptable (“it is the way it should be”). Different evaluations of our respondents were caused by the fact that other and other factors were coming into their minds. The resulting evaluation of respondents was influenced by the relevance the respondents assigned to the factors, or the number of factors was crucial.

The process of evaluation of examples was identical in case of Czech and Swedish respondents. They created their definitive evaluations the same way. They reported the factors on the basis of which they decided. Whether or not the example was acceptable for them was decided by the factors they took into consideration.

Legitimate Intertextuality

Both Czech and Swedish respondents characterized the legitimate intertextuality rather in negative terms. In the scope of both research the respondents commented rather on phenomena which were missing in the examples (what was missing in the examples) than on phenomena which were acceptable. As in the case of Swedish research, where the respondents identified a problem, there was more space for discussion. In the case of some Czech respondents the identification of the problem, or problem-free evaluation were preceded by a phase of more general description of a student text.

The Swedish research identified two factors the respondents used for defining the boundaries between acceptability and unacceptability:

- the scope of information the student transferred from the source text into his or her text “with his own words”,
- the presence of the reference to the source material.

Both of these factors were also recorded in our research, where we noticed other factors as well:

- the accuracy of transferred information and their significance for work (*the presence of secondary reference*—the presence of the reference to the source, from which the source material was derived and which the student based his or her work on; *change of the context of the transferred information*; *simplification of transferred information*),
- the benefit of the work.

If the factor occurred in case of more than one respondent, then the respondents characterized it consistently. This was not recorded in the Swedish study. Similarly as in the case of Swedish research, the failure in fulfilling of one of the factors resulted in illegitimacy of the text.

It can be finally stated, similar to the case of the Swedish research, that our participants believed that it was not necessary to refer to the source when transferring generally known information. The conformity between Czech and Swedish results was in conflict views affecting the type of information which may be considered well-known and its length. We detected conformity also in opinions of what can be regarded as an appropriate intertextuality. So the appropriate intertextuality may include the referred parts (such as the quotations or paraphrases), as well as non-referred parts in the form of transfer of generally known information.

Some Swedish respondents said that they found legitimate intertextuality even in the case when the student is copying from the source (the copying technique). They justified their opinion by the fact that it is connected with the learning process. We did not notice these opinions in the Czech study.

In case of the Czech respondents we detected a phenomenon the Swedish study did not identify. Some respondents said that the submitted example was acceptable. Further they claimed that there would be an option to increase the correctness of the text, if a student added more information into the text. By adding such pieces of information, a student shows that he or she has mastered certain academic skills and knowledge. At the same time, inclusion of such information in the text enables both the reader and the evaluator to work in depth with the information included in the text (for example pagination stated next to the reference to a paraphrased text enables easier tracing of the information in the source text of a student work).

Illegitimate Intertextuality

Both Swedish as well as Czech participants identified and described number of examples which were evaluated as inappropriate intertextuality. In accordance with the Swedish research we can say that the respondents, even though they evaluated such cases as inappropriate, mainly explained this situation in a different way than by trying to deceive intentionally.

In accordance with the Swedish research we found out that our respondents identified a certain form of illegitimate intertextuality which resulted from various reasons as in case of the foreign study. It should be noted that we recognized the same possible causes: the lack of knowledge of the rules for the recognition of the authorship and of the rules for the use of the source; and the process of controlling academic skills in the field of academic writing.

Both Swedish and Czech participants evaluated and explained intertextual links in different ways. They also sometimes explained the illegitimate intertextuality in variable ways. In contrast to the Swedish respondents, our respondents evaluated plagiarism without considering the subjective aspect of behaviour (meaning their intention to plagiarise). In contrast to the Swedish respondents, our respondents identified the case as plagiarism only if the criteria which they themselves attributed to plagiarism had been fulfilled.

Restrictions in the Evaluation of Legitimate or Illegitimate Intertextuality

Restrictions in the evaluation of intertextuality were not found in the Swedish study.

The fact that the academic is not an expert in the disciplinary field of the student work may significantly reduce the possibility to competently assess whether or not the intertextuality is legitimate or illegitimate in a particular case. Academic staff, even though they are not experts in the given field, also evaluated the field of transfer and the reduction of data from the source material as if they were judging a work from their own specialization. They are, however, aware of the fact that they may not be competent enough to assess this part affecting this type of intertextuality from information transfer point of view. Their evaluation in this field reflects whether the transmitted information has identical content. They are able to identify a change in the meaning of transferred data. However, they are not able to assess the text in relation to its specific focus.

Conclusion

This study pointed out different evaluations of legitimate and illegitimate intertextuality in student works by academics. In these cases the respondent bases the assessment not only on the text of a student work, but also with reference to the source from which the information was taken.

We found out that both the Czech academics as well as the Swedish respondents conflicted in their opinions on what constitutes legitimate and illegitimate intertextuality. In case of our respondents. However, we noticed greater differences not only in the case of assessment of acceptability or unacceptability of the intertextual links, but also in the case of assessment concerning the question whether or not it is the case of plagiarism.

Czech and Swedish respondents evaluate intertextuality in a capable manner. Swedish respondents evaluated intertextuality on the basis of the factors used by Czech academics as well. Czech respondents used also different factors for their assessments.

Opinion consensus has also been detected in the case of the factors that characterized the legitimate intertextuality:

- referred parts of the text (paraphrases and quotations),
- non-referred parts of the text (transmission of generally known information).

Czech respondents determined cases of plagiarism to a greater extent than Swedish academics. Swedish respondents identified the case as plagiarism only if they had

exhausted all the possibilities suitable for the explanation of the case. On the other hand, Czech respondents identified the case as plagiarism if all their criteria for accusing the case of plagiarism were fulfilled. This difference also explains the conflict in the assessment of plagiarism between Czech and Swedish respondents.

We would like to conclude this study with a proposal of recommendation directed to the educational practice and university administration.

The presented study is to a large extent separated from the practice, because it shows the evaluation only in the conditions when the respondent has the source material available. The pieces of information are transmitted from this source material to a student work, and the respondent knows the localization of transmitted information in the source text. In practice such a case of evaluation can be the assessment of outputs from the detection system. This study examines how different academics may evaluate identical output from the detection system in different ways. Therefore, it is appropriate that more academics should be involved in the evaluation (for example the head of work, opponent, or the members of the examination committee).

We found out that if a student work is assessed by the academic whose focus does not match the field within which the work is processed, he or she can suitably assess not only the formal aspect of work, but also, if the source material is available, he or she can assess the legitimacy of the transfer of information. These academics are aware of the fact that they are not certain of considering the informational aspect of work. Therefore, it is possible to state that academic staff from another field than in which the presented work is, are able to competently assess the formal aspect of work, and, in case of the outputs from the detection system, they are able to competently assess legitimacy of the transfer of information between the source and a student work. However, they will not be able to evaluate the professional aspect of work.

Czech academics also mentioned a series of recommendations how to add to the text. It will make the work with the text easier not only for a reader, but also for an evaluator of the work. These are practical observations resulting from the experience of academic staff working with the resources. It is appropriate that the students are familiar with these requirements:

- indicating pagination of a paraphrased text,
- in case of the quotations of foreign sources, the original text should be indicated in the footnote apparatus of work,
- indicating information on secondary sources (the sources used in the text from which we derived some information) in the text.

The study continues to be carried out. To compare the results with the Swedish research, only first eight respondents were chosen. Our aim is to continue in the study until the interviews still enrich the variants of factors identified up to now, and which were used for the evaluation. We have currently 16 interviews available, and we must state that we still notice different views of how to assess intertextual relations in student works.

Appendix (Pecorari and Shaw, 2012, p. 163–164)

Is this an example of a good way to use and refer to sources?

Why/why not?

(If the answer was that it is problematic, the following questions were asked as well.)

What would you do about it if you found it?

How would you explain to the student that this was a problem?

How would you tell the student what to do to fix it?

Would you take any actions against the student?

Probe specifically about disciplinary action.

Is this plagiarism?

Why/why not?

After all the passages were looked at:

If you imagine a continuum between fully appropriate and entirely unacceptable, where would you place these examples?

What is it that makes number x more/less acceptable than number y?

And follow up with other comparisons.

Probe for these if they don't emerge:

- quantity in one passage
- quantity overall
- what section it's in
- only so many ways to say the same thing
- the fact of writing in a second language
- the fact of being new to academic discourse

Now you've heard what I'm interested in, is there anything I haven't asked but should have?

Do you have any questions for me?

References

- Borg, E. (2009). Local plagiarism. *Assessment & Evaluation in Highe Education*, 34 (4), 415–426
- East, J. (2010). Judging plagiarism: a problem of morality and convention. *Higher Education*, 59 (1), 69–83
- Flint, A., Clegg, S., Macdonald, R. (2006) Exploring staff perceptions of student plagiarism. *Journal of Further and Higher Education*, 30 (2), 145–156
- Jager de, K. & Brown, C. (2010) The tangled web: investigating academics' views of plagiarism at the University of Cape Town. *Studies in Higher Education*, 35 (5), 513–528
- Pecorari, D. & Shaw, P. (2010) University teachers discussing plagiarism: divided perspectives on teaching writing and shaping a culture of honesty. *4th International Conference, 21–23 June 2010, Newcastle Upon Tyne, UK*
- Pecorari, D. & Shaw, P. (2012) Types of student intertextuality and faculty attitudes. *Journal of Second Language Writing*, 21 (1), 149–164
- Roig, M. (2001) Plagiarism and Paraphrasing Criteria of College and University Professors. *Ethic and Behavior*, 11 (3), 307–323

Author

Blanka Farková, blanka.farkova@seznam.cz, Jan Amos Komensky University Prague, Czech Republic

Copyright © 2013 Blanka Farková: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

A CULTURAL BRIDGE FOR THE ACADEMIC CONCEPT OF PLAGIARISM: A COMPARISON OF CHINESE AND BRITISH CULTURAL CONCEPTS OF PLAGIARISM BY CHINESE MASTER'S GRADUATES OF UK INSTITUTIONS EMPLOYED BY SINO-FOREIGN JOINT VENTURES IN SHANGHAI, CHINA

Stephen Gow

Abstract: Cultural and developmental perspectives (Flowerdew & Li, 2007) of plagiarism are explored through interviews with Chinese graduates of UK Master's degrees after they have returned to work in China. Plagiarism in the context of the participants' educational history, life in the UK and their return to China were investigated. These accounts provided narratives of their development of the concept of plagiarism and a cultural comparison of the British and Chinese cultural understandings of the term. Narratives were analysed for the use of interpretive repertoires (Gilbert and Mulkay, 1984) to identify the significant commonalities and inconsistencies within and between the participants' accounts. The findings suggest that the participants use UK institutional vocabulary for plagiarism which they have developed on their Master's course and in subsequent work. Furthermore, participants insist that the cultural concepts of plagiarism in China and the UK are equivalents, the only difference being the strictness with which rules are applied in the UK system. In contrast, their accounts of Chinese pre-university and undergraduate education reveal how educational expectations shape a divergent Chinese concept of plagiarism. Finally, the potential for these returning Chinese graduates to act as a cultural bridge for a rigorous application of plagiarism regulations within TNHE education is discussed.

Introduction

International education has faced a perceived plagiarism “epidemic” (Howard, 2004) as educators attempt to deal with the shift to the era of the Internet and globalisation (Sutherland-Smith, 2008). In this context, Chinese education has been singled out for particular attention as a result of issues concerning academic integrity and the cross-cultural conceptualisation of plagiarism. This attention is a symptom of the success and side effect of China's ambitious educational reform and internationalisation. In addition to being the world's largest education market (Wang, 2009), and leading source of international students (Counsell, 2011), China's scientific publication output has overtaken the UK and reached second place behind the US (Clarke and Plume, 2011). This has led to the comparison of the Western (mainly English speaking countries: UK, USA, Canada, Australia, New Zealand) and Eastern (particularly Confucian Heritage Cultures: China, Japan, Singapore, Vietnam) philosophical, legal and educational roots of plagiarism, and raised questions about the future form of research in a globalized world.

This paper aims to explore the intercultural (Gu, 2009) nature of plagiarism through examining a unique sample of the Chinese population who hold UK master's degrees and are employed in transnational higher education (TNHE) in China. Through in-depth interviews, the study creates a narrative which can be analysed for use of interpretive repertoires (Gilbert and Mulkay, 1984). The analysis will be used to explore the cultural and developmental hypotheses of plagiarism research in a new context. The experiences of this sample are by no means representative of the Chinese population. However, the participants' knowledge and experience of both UK and Chinese education provide an opportunity for an informed comparison of the Chinese and UK concept of plagiarism and reflection on students' development of this concept.

Background

The background of intercultural plagiarism research is highly complex and controversial. See Sutherland-Smith (2008) and Bloch (2012) for historical overviews.

Chinese Context

The development of the Chinese term for plagiarism; *chaoxi* (抄袭/to copy or steal) (also *piaoqie* 剽窃/to steal writing) (Hu and Lei, 2012), has occurred within the bounds of Confucian heritage culture (CHC). *Chaoxi* has not developed in the context of economic, educational and scientific development as it has in the West due to a Confucian opposition to profiteering, the legal system and didactic approach to education (Alford, 1997). Historically, education in China was dominated by imperial examinations, the *keju* (科举, until 1905), which required memorisation and analysis of the four books and five classics of Confucian literature (Miyazaki, 1981). As a result, citation was unnecessary and would even be interpreted as insulting to the learned reader (You, 2010). This educational culture has permeated into the modern era. The National Higher Education Entrance Examination, known as the *gaokao* (高考) has provided equality and quality in Chinese secondary education since the 1978 post-Mao reforms. However, this has relied on the traditional CHC didactic methods encouraging memorization of declarative knowledge rather than critically assessing the multiple perspectives of sources (Gao, 2012). These techniques are then continued at university level with little attention to intertextuality or attributive source training (Hu and Lei, 2012).

This educational background has been exacerbated by the successful yet breakneck pace of development and internationalization of HE in China. The consequences of this have been the unattributed translation of foreign research articles into Chinese (Li and Xiong, 1996) and a lack of originality in Chinese research (Ye, 2007 in Liu, 2009). In recent years, higher publication targets and policy driven targets in Chinese universities have led to a "publish or perish" culture which in turn has led to plagiarism and rushed research (Yi, 2011:502). As a result of this intense educational pressure, academic corruption and plagiarism have been reported in the national and international media (Baty, 2009, Guo, 2010, Osnons, 2010). Recent research emerging from China has shown the extent and proposed solutions to this problem (Zhang et al., 2010, Yi, 2011). In reaction, the government's desire to crack down "mercilessly" on corruption

and to revamp postgraduate education, placing “a premium on integrating learning with thinking” has also been clearly stated in its 2010–2020 Educational reform plans (Educational-reform, 2010:25 & 50).

21st Century Paradigm

Since the turn of the century, universities around the world have experienced a cultural shift in the approach to plagiarism due to internationalisation and the Internet. Cut and paste functions, and the explosion of availability of online sources have led to a noticeable increase in plagiarism (McCabe, 2005). Kress (1998) and Howard (Howard, 1999) note that as a result the internet is a contested intertextual space which challenges the traditional romantic notions of authorship. This situation has been magnified by the increase of international students on campuses, some of who have limited experience of non-exam based assessment, source use and contrasting approaches to textual practices (Carroll, 2008). These changes have made an explicit pedagogical and procedural approach a necessity (Stefani and Carroll, 2001). Whether it be Howard's (also Howard, 1999, Pecorari, 2010) distinction between blatant plagiarism and patchwriting or Chandrasoma, Thompson and Pennycook's (2004) transgressive and non-transgressive intertextuality, the aim is to define the concept within the global, post-internet paradigm. Consequently, in order to maintain quality academic standards and integrity (Carroll and Appleton, 2001) UK universities have moved away from plagiarism as an academic death penalty (Howard, 1995), to more consistent and measured punishments which take into account intention and the cultural complexity of the concept (Park, 2004). In addition, universities have adopted an accommodation model, providing support for changing student demands, including explicit instruction of avoiding plagiarism (Ryan and Carroll, 2005). While the development of software has successfully assisted the detection of non-originality for tutors and students, it seems to handicap International students (Zobel and Hamilton, 2002) providing further complexity to the issues of writer development, intent to deceive and the cultural concept of plagiarism.

UK's China problem?

UK higher education's (UKHE) reputation for quality education, as reflected in world rankings, has attracted increasing numbers of Mainland Chinese students (MCS) (Counsell, 2011). They are a major revenue stream for UKHE, especially since the signing of bilateral government agreements in 1999 (Gu and Brooks, 2008). In addition, they are a source of unique global research perspectives (Rastall, 2009), which in turn improve rankings. Accommodating these global perspectives of this significant minority on British campuses (UKCISA, 2012), provides a challenge to UKHE.

The UK education system prioritises the dissemination of knowledge through critical thinking and essay writing (Durkin, 2008) with roots in the Socratic principal of the extension of knowledge via inquisition (Tweed and Lehman, 2003). Due to MCSs' educational background, especially their *gaokao* experience, the development of independent, critical thinking skills (De Vita and Bernard, 2011) and adoption of Western essay writing practices is a significant adjustment (Gu, 2009). Smith and Zhou (2009)

found the following image of Chinese students emerged from interviews with British staff; on the one hand polite, hard-working and respectful, on the other rote learners with a lack of academic curiosity and an unwillingness to mix with other students. Unfortunately, the positive traits of Chinese students have been somewhat overlooked due to staff having to spend more time dealing with the strugglers. In addition to the linguistic difficulties faced by non-native English speakers (NNES), cultural differences have been magnified by the significant presence of Chinese International students on UK campuses (Sowden, 2005). This has resulted in the perception of a “Chinese problem” on UK campuses (Smith & Zhou, 2009:133) and as a cumulative result Chinese students have gained a reputation for plagiarism (Gu & Brooks, 2008).

The stereotype of the Chinese rote learner and plagiarist is unfortunate. As Gu and Brooks (2008) note, Chinese students are successful in UKHE and a number of Chinese students suit the creative and methodical educational approach abroad. This has resulted in a brain drain of Chinese talent (Huang, 2003) with only a quarter of those studying abroad returning between 1987 and 2005 (Mohrman, 2008). The government programmes, such as the 2008 Thousand Talents (千人计划) programme (Yi, 2011) aim to attract returnees, often referred to as *haigui* (海归, sea turtles) (Wattanavitukul, 2002).

Theory

Ten perspectives have been identified by Flowerdew and Li (2007) within plagiarism research, of which two are especially significant for Chinese students and this study:

- The cultural perspective accepts that cultures are different and concentrates on studying Chinese perceptions of plagiarism, usually within China (Matalene, 1985, Pennycook, 1996, Shi, 2004, Shei, 2005, Valentine, 2006, You, 2010, Hu and Lei, 2012)
- The developmental perspective acknowledges that cultures are not static and that perceptions and interpretations of plagiarism will change as a result of internationalization (Flowerdew and Li, 2007). In addition, Chinese students must develop to bridge the gap between cultures, which Durkin (2008) names a “middle way”, in order successfully graduate from Western institutions (Gu and Brooks, 2008, Gu, 2009)

The cultural theory in respect to Chinese culture and plagiarism has developed since the 1980's (Matalene, 1985). There were especially significant developments from scholars based in Hong Kong in 1990's, such as Deckert (1993), Scollon (1995) and Pennycook (1996). With increased internationalization of universities however, these cultural examinations produced a “crude East/West dichotomy” (Pennycook, 1996) which created Chinese stereotypes when studying abroad (see Liu, 2005, Sowden, 2005). As a result the developmental approach proved to be a more constructive theory to examine students and the concept of plagiarism in an international context (Flowerdew and Li, 2007).

This research project shall examine plagiarism through the cultural and developmental perspectives of Chinese graduates who have studied in China at secondary

and undergraduate level, and then successfully gained a UK Master's degree before returning to work in higher education in China. This aims to take a long term view of plagiarism as not merely a concept faced by students in one particular environment but as a developing cultural concept throughout their educational careers.

Sino-Foreign Joint Ventures

A joint venture (JV) is a partnership between a receiving country (developing nation, e.g. China) and a providing country (developed nation e.g. UK) (Rastall, 2009). The providing "parent" university opens a branch campus in the receiving country to offer the "parent" university's degrees or a joint degree between the parent and partner institution (Cao, 2011). Since the 2003 Sino-Foreign Higher Education Law a number of high profile JVs have opened in China (Drew and Britain, 2008). The Chinese staff require international experience and qualifications in order to work in the intercultural environment.

Research Focus & Hypothesis

The main focus of the research reported in this paper is to:

- Examine participants' accounts of plagiarism and the relevance of this cultural concept to their educational, professional and social development.
- Compare and contrast their experience of plagiarism in Chinese and British educational contexts.

The research aims to extend the developmental and cultural perspectives to the new context of returning Chinese graduates from UK institutions. The research question is therefore:

- To what extent do these returnee scholars understand and transmit the concept of plagiarism when returning to work in China?

The hypotheses are as follows:

- The participants will describe a difference in the cultural concept of plagiarism in China and the UK and the continuing process of development which they have undergone to adapt to the culturally varied concept of plagiarism.
- The participants' accounts will display how Chinese returnees with Master's degrees from UK institutions develop an intercultural understanding of plagiarism and act as a cultural bridge for the development of the concept.

Methodology

Participants

The sample includes teachers from English for academic purposes (EAP), business, engineering and management, and a number of participants from administrative roles and student support. The researcher gained access to five participants from each of the JVs (total 10), selected on the basis of Miles and Huberman's (1994, p. 28) criterion sampling:

- Be a Chinese national;
- Have studied at secondary and tertiary undergraduate level in a Chinese higher education institution;
- Have completed a Master's degree at a UK higher education institution;
- Have returned to live in China and work in a Sino-foreign JV institution in Shanghai.

Interviews

This study uses Holstein and Gubrium's (1995) concept of the active interview. Kvale (1996) notes that interviews engage the ambiguous and contradictory nature of participants' experiences while providing openness to new insights into phenomena. Due to the illicit (Cohen *et al.*, 2007) and cross-cultural nature of the topic, a familiarity with Chinese interactional dynamics (Shah, 2004) was significant in establishing a rapport conducive to in-depth discussion (Gubrium and Holstein, 2002). Having lived in China for five years, working in JVs and national institutions, my familiarity with the context helped to carry out effective interviews. The English language (with Mandarin used to clarify terms) interviews (approx. 1 hour) took place at private locations in the JV institutions between April/May 2012. They were recorded and later transcribed in a simple format suited to phrasal and sentence level analysis (Gilbert and Mulkay, 1984).

Open questions were used for the exploration of the sensitive issue of plagiarism. The interview was divided into three stages which create a narrative of the participant's development through his or her time studying in the UK and returning to China:

1. Educational Background
 - University and subject of undergraduate study in China;
 - University and subject of postgraduate study in the UK;
 - Reasons for studying abroad.
2. Living & Studying in the UK
 - Problems faced living and studying in the UK;
 - Encounters with plagiarism and a definition;
 - Informal and formal instruction on plagiarism.
3. Working & Living back in China
 - His or her current work role and the relevance of his or her UK qualification to securing the position;
 - His or her encounters with plagiarism in current work role;
 - His or her perceived relevance of plagiarism to the broader context of originality in China;
 - Comparing the British and Chinese concepts of plagiarism;

Ethics

Plagiarism is an illicit activity synonymous with cheating (Hayes and Introna, 2005) as it concerns possible "guilty knowledge" (De Laine, 2000, p. 85). The research, therefore, may reveal embarrassing or damaging information concerning individuals

or institutions, such as criticism or accusations (Stanley and Wise, 2010). As a result, extra care was taken to explain the context of the research and aims via the *Research Information* and *Informed Consent* forms. The participants were given written and verbal reassurance of their right to withdraw at any time. The anonymity of all people involved and institutions has been maintained through the removal of identifiers. Numbers have been used (e.g. P1, P2, etc.) to reduce the risk of coincidental confusion of participants with other identities. The data was kept on password protected USBs and on in a back-up hard drive, to which only the researcher had access.

Interpretive Repertoires

Interpretive repertoires are a method of discourse analysis based on ethnomethodology which identifies global patterns in accounts of actions and beliefs (Potter and Wetherell, 1987). First used by Gilbert and Mulkay (1984) to analyse discrepancies between biochemists' *empiricist* and *contingent* repertoires, the authors found that scientists required two repertoires to make sense of their empirical findings. This method is especially suited to exploring controversial and sensitive issues (Wetherell, 2005) such as plagiarism, as it explores representations of "recognizable themes, common places and tropes" (Potter and Wetherell, 1987) known as *doxa* in discourse (Barthes, 1977 in Wetherell, 1998). The repertoires are identified by the repetition of exact matches of phrases or similar descriptions embedded in the historical context (Edley, 2001). The commonalities and inconsistencies between participants and repertoires emerge through continuous, in-depth involvement in the design, interviews, transcription and reevaluation of the data (Edley, 2001).

Findings

The analysis of the participants' discourse reveals five repertoires through which a sense of plagiarism and study in the UK is provided. The following section will include a brief description of these repertoires and some key examples.

Comparing Plagiarism & Chaoxi

When asked directly to compare plagiarism and *chaoxi* (抄袭/to plagiarise), the common translation, the participants repeated the notion that the "interpretation" was the same yet the "methods" were different. However, after closer questioning the participants began to describe the different educational contexts within which the term is used. Due to both their educational background in the UK and working in the JV, they show a familiar usage of institutional vocabulary referring to plagiarism as an "academic crime" (P3, P10) or "academic dishonesty" (P2). In contrast, their Chinese experience with *chaoxi*, revealed a "lighter sense" of the phrase. This referred to more general cheating in exams and homework starting in primary school, rather than in the academic sense. There was a strong sense of "sharing" or "helping" fellow students rather than cheating, which at times received "silent approval" (P3) from teachers. As one participant mentioned, the teacher would "close one eye, open one eye" (P4).

Once reaching undergraduate level in China, *chaoxi* would be mentioned but formal instruction on the definition and avoidance of misconduct would not be offered.

Assessment and Referencing

The participants frequently refer to the difference in assessments, particularly writing assignments. In the Chinese context, it seems that in many of cases *chaoxi* or cheating is referred to in respect to exams. When written assignments were requested, such as dissertations, the participants referred to them being similar to a literature review without major referencing; this was a quicker and easier process than in the UK (P1, P5). One participant had never written an essay prior to studying in the UK and thus felt at a disadvantage to start with (P7). For more of the participants, the referencing system proved to be a confusing and time consuming task. In relation to paraphrasing and quotation, a minimal amount was required in China; however it was not strictly enforced. The participants admitted to struggling or failing early assignments (P1, P5, P7, P10) or referred to having acquaintances who were accused of plagiarism (P1, P2, P3, P4, P8, P9), similar to findings by Ryan and Carroll (2005). Furthermore, the timing and requirements of the UK MA assignments confused the participants. In one case a participant was shocked to have an exam after three months, claiming she had not learned anything. This was reinforced by P1, who was surprised to be asked to brainstorm a topic about which she knew nothing, noting “you give me input and I’ll give you output”.

Independence, Responsibility & Multiple Perspectives

Independence dominated not just the participants’ academic life, but also their social life in the UK. In China they noted that social activities, food, accommodation, opening bank accounts were all supported or clearly explained by a representative of their institution or fellow students. Many participants noted that they learned to cook in the UK out of necessity and also spent their first month busy, lonely and confused. There was a common view that “nobody will tell you anything” (P1, P4, P8, P10) and the responsibility being on the student to read the relevant documents and seek help themselves, for example at the English language centre. This was reflected in the academic experience. In China “they will tell you everything...everything is in the book” (P1), this was repeated in one form or another by most participants. While comforting, this did not breed a sense of responsibility or choice in the educational process. Students would have little academic choice and considerably more classroom hours, including compulsory political courses. It seems the responsibility lay more on the teachers to teach and take care of the students, who have little option to think or make choices. The contrast is clearly exemplified by P5:

If you are sleeping [during class] or anything, I think he [the monitor or tutor] will call my parents

In contrast to the UK:

You can come, or you can don’t come [sic]. It’s your personal responsibility.

As a consequence many participants noted, the focus of the repertoires shifts from *they*, referring to the Chinese teachers and staff, to *self* in the UK. For example “push myself” (P2) or “push yourself...you need to study for yourself” (P5). This shift to self-responsibility leads to new conflicting perspectives of truth and authority. One participant refers to the time consuming process of the “switch from Chinese thinking to English thinking” (P4), others refer to changing habits (P10) and getting used to British thinking (P1).

With this newfound responsibility, the participants had to adjust their perception of a teacher of a sole authority to a lecturer as a research guide. This shift was preferred by certain participants who positively note that the lecturers have office hours (P4, P5) and the value of one-on-one sessions (P4, P6, P10). The British lecturers were also more likely to relate information to the real world than to textbooks or exams. Contrastingly, two participants (P3 & P5) faced problems with busy dissertation advisors; this was a situation the students were responsible for resolving.

As a result, an independent, critical self-identity was created within the group of students. Although strange at first, the multiple, often conflicting perspectives in research, group and class work started to form a critical dimension to participants’ thinking (P1, P2, P3, P6, P7, P10) and also improved their communication skills (P1, P2, P3, P10). As opposed to communal studying in China, where the students studied together, in the UK the participants were able to appreciate different perspectives in reaching group objectives and assessing sources, a highly valued skill (P1, P2, P3, P6, P7).

Strictness & seriousness

The role of tutors and independence are contrasted with of strict and serious attention in matters of attribution of knowledge. In the British context, the participants describe both written and oral warnings about committing plagiarism. All participants made a distinct reference to this, here are two representative examples:

P2—Everything academically is very serious... Formality and seriousness. Strictness and formality... Academic seriousness... In the UK I found plagiarism is more strict.

As opposed to:

P5—Chinese won’t check it, time and money... maybe for postgrad they will check it... Even if you plagiarise you will be allowed to graduate. We do not focus on plagiarism according to Chinese culture... if no one finds it, it will be ok.

As a result of this strictness the participants focused on the mechanics of citation and avoiding plagiarism, with P5 mentioning “you just write the bibliography and quotations”. However, this simplistic description seems to overlook the process of adaptation to fundamentally different education system. Additionally, the role of the teaching staff provides a paradox. On the one hand the Chinese teachers are watchful over students at all times (P5) and timetables are full (P4, P10), yet on the other hand teachers are “too busy” (P1, P5), or even lazy and complicit (P4), in seeking out and finding plagiarism. This is contrasted with the relaxed or uncaring description of the British tutors who switch to enforcers at the sight of plagiarism.

Plagiarism in work and life

The UK qualification was seen as playing a major role in their ability to work in the JV institution. The participants felt that their experience in the UK had led them to be more open-minded and flexible (P10, P7) and that they benefited from research (P4), multi-tasking, communication, and group-work skills (P3). The role of plagiarism in the participants' jobs was more significant for English teaching staff and personal tutors than subject teachers and administrative staff. Poor language skills and the use of copy-paste were viewed as sources of student plagiarism which had been reduced by the use of software, such as *Turnitin* (P1, P3, P4, P6, P9, P10), which contrasted with the holistic, cultural experience of their own narratives.

A broader perception is of the difference in the standard of education provided by UK and Chinese institutions and the issue of academic corruption. A proportion of participants (P2, P3, P6, P10) reflected general concerns about academic plagiarism at Chinese universities. The "publish or perish" culture and political nature of universities (P6, P10) is identified as a cause, as are the lower standards of certain Chinese universities compared to the UK's trusted high standards (P2). These concerns are reflected in local media stories and also the government's educational reforms. These educational concerns were also reflected in a general concern in society for copyright issues. The participants saw the short-term benefits for economic development, such as counterfeit products and internet sites, such as Facebook compared to Weibo (P3, P4, P5, P7). Yet, there is a clear desire shown by the participants for China to innovate rather than follow in the future (P1, P3, P4, P6, P7, P8, P10).

Discussion

When asked directly on the difference between plagiarism and *chaoxi*, the participants implied the meaning was the same, however the action in the UK was more strict and serious. Analysing this repertoire in isolation from the previous repertoires would provide a clear case of the difference being the rules applied to plagiarism. However, simply concentrating on the rules and punishments overlooks the underlying approach to education in the two countries. In the Chinese context, the participants described instances of plagiarism involving copying from other students and also in the context of exams. In this sense, *chaoxi* resembles the dictionary definition of plagiarism i.e. "passing someone else work, idea off as your own" (O.E.D, 2013) rather than the academic definitions based upon new plagiarism pedagogy (e.g. Howard, 1999, Carroll and Appleton, 2001). In contrast, the British experience of plagiarism is imbedded in critical research essays showing awareness of intertextual and attributive practices. As a result, the whole academic process in the UK is encapsulated with the repertoire of independence. Self-responsibility and awareness develop through exposure to multiple perspectives.

Plagiarism represents the corruption of the process of independent and critical thinking which are the key to adding to the body of knowledge. Detailed citation and referencing not only avoids the corruption of knowledge through false attribution, but also provides links within the body of knowledge, evidence for assumptions and the logical formation of conclusions. The switch in the role of the UK tutor from a liberal

guide to a strict police officer in regards to plagiarism is contrasted with the Chinese teachers who reportedly were strict on attendance to class and conformity yet lenient on plagiarism. This is due to the different expectations of the systems. In the context of the Chinese exam-based educational experience, plagiarism is not strictly punished because the students' role is primarily to learn rather than to think and create. The concept of copying from a fellow classmate as plagiarism does not concern copying their way of thinking but strictly copying the correct answer; this bypasses effort rather than corrupts knowledge.

Conclusion

The participants in the study therefore do not return to China espousing an explicit difference between the British and Chinese concept of plagiarism. However this does not imply that they do not act as cultural bridge for the concept. The repertoires of strictness, independence, self-awareness, critical thinking and creativity, in addition to the perceived benefits of a UK education, such as collaborative and presentation skills, implicitly rely on the stricter definition of plagiarism. As Durkin (2008) suggests, the participants have adapted to a "middle way" between British and Chinese culture. Their attention to strictness is due to the traditional respect for authority; however in British higher education the authority is in the system rather than the teachers. Their strict adherence to the rules and the educational context has forced them to think critically and independently and within this process they have successfully adapted to the system. Therefore when they have returned to China and work in the JV, they present a positive example of the transformation to students. In instructing, offering advice and collaborating with Chinese and international staff and students, they implicitly reinforce the stricter academic concept of plagiarism implicitly and explicitly through the use of institutional vocabulary.

Implications

As more UK graduates return to live and work in China, their views will influence the country's future trajectory. These findings imply a strict enforcement of policies and rules regarding plagiarism in universities and the introduction of essay based qualitative assessment at a lower level may lead to the "integration of learning with thinking" (Educational-reform, 2010:25). However, as with Chinese economic reform, it may be a case of reform with Chinese characteristics, or a middle way (Durkin, 2008) to produce similar results. The participants saw the need for China to lead rather than follow in the future but plagiarism and copying may hinder national development in the long-run.

Previous research into Chinese students and plagiarism in the UK, particularly Gu and Brooks (2008) and Durkin (2008) is reinforced by this study. By placing development into a broader context, the Chinese graduates' pre- and post- UK educational and career history emerge. The accounts show that the measures taken to accommodate international students are successfully assisting Chinese learners to adapt however raises the question of how far to accommodate them without creating dependence. A deeper understanding of Chinese students' backgrounds and their intercultural

accounts of education can assist UK educators to refine the accommodation process and inform Chinese students intending to gain UK qualifications. Furthermore, the intercultural accounts provide a unique insight into the implicit role of the concept of plagiarism in educational culture in the era of the Internet and internationalisation.

Further Research

This project presented a small sample of UK educated Master's students from China working in one particular field. Further investigations could trace the alumni UKHE in the global employment market and differing reflections on the UK educational experience. In addition, the comparison of the academic integrity expectations at Master's level in China and the UK, as suggested by the quote "... maybe for postgrad they will check it..." (P5), has yet to be published in English. Citation analysis by Bloch and Chi (1995) and Shi's (2002) examination of publication practices by Western trained scholars in China has shed some light in the difference in publications, however establishing the differences in academic integrity instruction at equivalent degree levels would provide an interesting study. As I have now returned to the UK to complete a PhD, I intend to explore the developmental process of Mainland Chinese Master's students' academic writing with specific attention to academic integrity.

Reference List

- Alford, W. (1997). *To steal a book is an elegant offense: Intellectual property law in Chinese civilization*: Stanford University Press
- Baty, P. (2009). The Long March Times Higher Education
www.timeshighereducation.co.uk/story.asp?storycode=408794 [Accessed 10/02/13], 29/10/2009
- Bloch, J. (2012). *Plagiarism, Intellectual Property and the Teaching of L2 Writing*: Multilingual Matters Limited
- Bloch, J. & Chi, L. (1995). A comparison of the use of citations in Chinese and English academic discourse. *Academic writing in a second language: Essays on research and pedagogy*, 231–274
- Cao, Y. (2011). Branch Campuses in Asia and the Pacific: Definitions, Challenges and Strategies. *Comparative and International Higher Education*, vol. 3, 8–10.
- Carroll, J. (2008). Assessment Issues for International Students and for Teachers of International Students. *The Enhancing Series Case Studies: Internationalisation*
- Carroll, J. & Appleton, J. (2001). Plagiarism: a good practice guide. *Oxford Brookes University*
- Chandrasoma, R., Thompson, C. & Pennycook, A. (2004). Beyond plagiarism: Transgressive and nontransgressive intertextuality. *Journal of Language, Identity, and Education*, 3, 171–193
- Clarke, L. & Plume, A. (2011). *Knowledge, networks and nations: Global scientific collaboration in the 21st century*. London.
- Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education* London: Routledge
- Counsell, D. (2011). Chinese Students Abroad: Why They Choose the UK and How They See Their Future. *China: An International Journal*, 9, 48–71
- De Laine, M. (2000). *Fieldwork, participation and practice: Ethics and dilemmas in qualitative research*: Sage Publications Limited.
- De Vita, G. & Bernard, M.J. (2011). Fostering reflective learning in Confucian Heritage Culture environments. *Innovations in Education and Teaching International*, 48, 183–193
- Decker, GD. (1993). Perspectives on plagiarism from ESL students in Hong Kong. *Journal of Second Language Writing*, 2, 131–148

- Drew, S. & Britain, G. (2008). *Trans-national education and higher education institutions: Exploring patterns of HE institutional activity*: DIUS Publications
- Durkin, K. (2008). The middle way: East Asian master's students' perceptions of critical argumentation in UK universities. *Journal of Studies in International Education*, 12, 38–55
- Edley, N. (2001). Analysing masculinity: Interpretative repertoires, ideological dilemmas and subject positions. *Discourse as data: A guide for analysis*, 189–228
- Educational-Reform, (2010). The 2010 announcement of the 2010–2020 National Medium to Long Term Education Reform and Development Plan Outline (2010–2020)
https://www.aei.gov.au/news/newsarchive/2010/documents/china_education_reform_pdf.pdf [Accessed 10/02/13]. In M. O. Education (ed.)
- Flowerdew, J. & Li, Y. (2007). Plagiarism and second language writing in an electronic age. *Annual Review of Applied Linguistics*, 27, 161
- Gao, H. (2012). The Education System That Pulled China Up May Now Be Holding It Back *The Atlantic*, <http://www.theatlantic.com/international/archive/2012/06/the-education-system-that-pulled-china-up-may-now-be-holding-it-back/258787/> [Accessed 10/02/13]
- Gilbert, GN. & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse*: Cambridge University Press
- Gu, Q. (2009). Maturity and Interculturality: Chinese students' experiences in UK higher education. *European Journal of Education*, 44, 37–52
- Gu, Q. & Brooks, J. (2008). Beyond the accusation of plagiarism. *System*, 36, 337–352
- Gubrium, JF. & Holstein, JA. (2002). *Holstein. Handbook of Interview Research*: Thousand Oaks: Sage Publications
- Guo, JX. (2010). Academic corruption undermining higher education: Yau Shing-tung *China Daily, Hong Kong Focus (HK Edition)*
http://www.chinadaily.com.cn/hkedition/2010-06/02/content_9919871.htm [Accessed 10/02/13]
- Hayes, N. & Introna, L. (2005). Systems for the production of plagiarists? The implications arising from the use of plagiarism detection systems in UK universities for Asian learners. *Journal of Academic Ethics*, 3, 55–73
- Holstein, JA. & Gubrium, JF. (1995). *The active interview*: Sage Publications, Incorporated.
- Howard, RM. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57, 788–806
- Howard, RM. (1999). *Standing in the shadow of giants: Plagiarists, authors, collaborators*: Praeger Pub Text
- Howard, RM. (2004). *Plagiarism epidemics, media epidemics* [online].
<http://wrt-howard.syr.edu/Papers/Colby.htm> [Accessed 10/02/13]
- Hu, G. & Lei, J. (2012). Investigating Chinese University Students' Knowledge of and Attitudes Toward Plagiarism from an Integrated Perspective. *Language Learning*, 62, 813–850
- Huang, F. (2003). Policy and practice of the internationalization of higher education in China. *Journal of Studies in International Education*, 7, 225–240
- Kress, G. (1998). Visual and verbal modes of representation in electronically mediated communication: The potentials of new forms of text. *Page to screen: Taking literacy into the electronic era*. 53–79
- Kvale, S. (1996). *InterViews. An introduction to qualitative research writing*: Sage Publications, Thousand Oaks, CA
- Li, X. & Xiong, L. (1996). Scientific misconduct: Chinese researchers debate rash of plagiarism cases. *Science*, 274, 337–337
- Liu, D. (2005). Plagiarism in ESOL students: Is cultural conditioning truly the major culprit. *ELT journal*, 59, 234–241
- Liu, F. (2009). How is research on academic plagiarism in China conducted? A preliminary investigation of the recent change in the style of writing in an academic journal. *Educational Integrity: Creating an Inclusive Approach*. University of Wollongong, NSW, Australia

- Matalene, C. (1985). Contrastive rhetoric: An American writing teacher in China. *College English*, 47, 789–808
- Mccabe, DL. (2005). Cheating among college and university students: A North American perspective. *International Journal for Educational Integrity*, 1, 10–11
- Miles, MB. & Huberman, AM. (1994). *Qualitative data analysis: An expanded sourcebook*: Sage Publications, Incorporated
- Miyazaki, I. (1981). *China's examination hell: the Civil Service examinations of Imperial China* New Haven, London: Yale University Press
- Mohrman, K. (2008). The emerging global model with Chinese characteristics. *Higher Education Policy*, 21, 29–48
- O.E.D (2013) Oxford English Dictionary Online Edition www.oed.com/ [Accessed 10/02/13]
- Osnons, E. (2010). Science Cop Mugged. *The New Yorker*, 31/08/2010
- Park, C. (2004). Rebels without a clause: towards an institutional framework for dealing with plagiarism by students. *Journal of Further and Higher Education*, 28, 291–306
- Pecorari, D. (2010). *Academic writing and plagiarism: A linguistic analysis*: Continuum
- Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL quarterly*, 30, 201–230
- Potter, J. & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour*: Sage Publications Limited
- Rastall, P. (2009). Cooperation with Chinese Universities-Issues and Trends. In T. Coverdale-Jones & P. Rastall (eds.) *Internationalising the university: the Chinese context*. 57–71
- Ryan, J. & Carroll, J. (2005). Canaries in the Coalmine. In J. Carroll & J. Ryan (eds.) *Teaching International Students. Improving Learning for All*. 1–10
- Scollon, R. (1995). Plagiarism and Ideology: Identity in Intercultural Discourse. *Language in Society*, 24, 1–28
- Shah, S. (2004). The researcher/interviewer in intercultural context: a social intruder! *British Educational Research Journal*, 30, 549–575
- Shei, C. (2005). Plagiarism, Chinese learners and Western convention. *Taiwan Journal of TESOL*, 2, 97–113
- Shi, L. (2002). How Western Trained Chinese TESOL Professionals Publish in Their Home Environment*. *TESOL quarterly*, 36, 625–634
- Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21, 171–200
- Smith, PV. & Zhou, X. (2009). The dissonance between insider and outsider perspectives of the “Chinese problem”: Implications for one of the UK's largest undergraduate programmes. In T. Coverdale-Jones & P. Rastall (eds.) *Internationalising the university: the Chinese context*. Palgrave Macmillan: England
- Sowden, C. (2005). Plagiarism and the culture of multilingual students in higher education abroad. *ELT Journal*, 59, 226–233
- Stanley, L. & Wise, S. (2010). The ESRC's 2010 Framework for Research Ethics: Fit for research purpose? *Sociological Research Online*, 15, 12
- Stefani, L. & Carroll, J. (2001). *A briefing on plagiarism*: Learning and Teaching Support Network Generic Centre Assessment Series
- Sutherland-Smith, W. (2008). *Plagiarism, the Internet, and student learning: improving academic integrity* New York; London: Routledge
- Tweed, RG. & Lehman, DR. (2003). Confucian and Socratic learning. *The American Psychologist*, 58, pp. 148–149
- Ukcisa (2012). *International Students in UK Higher Education: Key Statistics* [online]. http://www.ukcisa.org.uk/about/statistics_he.php#table8 [Accessed 03/02/13]
- Valentine, K. (2006). Plagiarism as literacy practice: Recognizing and rethinking ethical binaries. *College Composition and Communication*, 89–109

- Wang, J. (2009). The Internationalisation of Higher Education in China: The Case of One University. *Internationalising the University: The Chinese Context*, 57–71
- Wattanavitukul, P. (2002). Hai Gui: The Sea Turtles come marching home in China's Awakening Dragon. [online]. <http://www.apmforum.com/columns/china19.htm> [Accessed 10/02/13]
- Wetherell, M. (1998). Positioning and interpretative repertoires: Conversation analysis and post-structuralism in dialogue. *Discourse & Society*, 9, 387–412
- Yi, L. (2011). Auditing Chinese higher education? The perspectives of returnee scholars in an elite university. *International Journal of Educational Development*, 31, 505–514
- You, X. (2010). *Writing in the devil's tongue: A history of English composition in China*: Southern Illinois University Press
- Zhang, X., Sun, H. & Zhang, F. (2010). Preventing Plagiarism and Academic Misconduct: A Case Study of Chinese Universities. *4th Annual Plagiarism Conference, 2010*. Gateshead, UK
- Zobel, J. & Hamilton, M. (2002). Managing student plagiarism in large academic departments. *Australian universities review*

Acknowledgement

The author would like to thank: Sanna Eriksson for her critical insights into my work and patience. I'd like to thank my family for providing me with emotional, intellectual and financial support. The participants in this project not only provided rich data but have made the research process fun and I look forward to seeing them again in the future. This article is adapted from my Master's dissertation submitted at the Institute of Education (IOE), University of London, September 2012, the full version of which is available on request. I would like to extend my gratitude to my dissertation supervisor, Dr. Caroline Pelletier and, all the members of the IOE's international programme team, particularly Dr. Will Gibson and Dr. Jude Fransman. Also, a further thanks goes to my current PhD supervisor Dr. Danijela Trenkic at the University of York for her support with drafting this article and furthering my research into new areas. Not forgetting the authors included in my reference list and all those who I have read over the years which have played a major role in shaping my ideas. Finally, I would like to thank my colleagues and friends who have shown an interest in and contributed to my research.

Author

Stephen Gow, sjg526@york.ac.uk, Department of Education, University of York, UK

Copyright © 2013 Stephen Gow: The author grants to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the author.

SLOPPY REFERENCING AND PLAGIARISM IN STUDENTS' THESES

Erja Moore

Abstract: The discussion about plagiarism in Finnish higher education is scarce. The statistics provided by National Advisory Board on Research Integrity (2012) revealed only two cases of plagiarism in Finland in 2011. This is in sharp contrast with international figures of plagiarism. This paper introduces a study about referencing practices in electronically published theses. The data consist of a sample of theses published in Theseus data base (an electronic publication forum for universities of applied sciences in Finland). Altogether 91 randomly chosen theses in the area of health and business studies were analysed. Most students master referencing and follow the good writing practice. However, a significant amount of theses lack in referencing. Examples of writing malpractice are diverse. Quotations are presented without quotation marks, the primary source is not identified, and referencing can be misleading and unclear. Several plagiarised text columns were also found: text from existing source is taken, sentences might be slightly altered, and the sources of the original text are copied. Based on the data, there is evidence that plagiarism is present in Finnish students' theses, and the writing instructions given by each HE institution are not always followed. It can also be concluded that local understandings of plagiarism and the criteria for accepted writing practice vary.

Introduction

Studying in higher education has changed. In educational discourse there has been a shift from teacher centeredness to student centeredness. Earlier, till the last decades of last century, studying meant attending lectures and taking exams, but along the change in the conception of learning students are supposed to learn more independently today. Barr and Tagg (1995, p. 13) defined this change in higher education as a change from 'instruction paradigm' to 'learning paradigm'. Under the instruction paradigm the main activity of higher education institutions was to arrange lectures. The learning paradigm ended "*lecture's privileged position*", "*college's purpose is not to transfer knowledge*", but to "*bring students to discover and construct knowledge for themselves*" (Barr & Tagg 1995, p. 14–15). Both teachers and students in higher education today seem to share the view that taking an exam shows only superficial knowledge and memorizing. Written assignments, on the other hand, are considered to indicate the student's actual learning and knowledge. While writing, it is believed that students gain broader knowledge as they also have to process the information they have studied.

It is a world-wide assumption that students learn at least the basics of academic writing before graduating from university. As Masters (2005, p. 282) puts it: "*In academic writing, the need to quote and reference accurately is a given. To ensure accuracy, reference systems are designed in great detail (...)*". Among the most used referencing systems are the APA style provided originally by American Psychological Association (APAstyle.org 2013), The Harvard Referencing system (see e.g. University of Exeter 2013) and MLA Referencing system provided by Modern Languages Association (see e.g. University of York 2013). The main idea in a referencing system is that the

author identifies the original author of the source and refers to that author when presenting text (citation) or content (paraphrasing) from that source.

Writing itself has completely changed in the last 10–15 years. Teachers in higher education took their initial studies in a completely different environment before the development of ICT. Writing meant taking a pen in your hand and moving it on paper. After you had made the plan of your essay or thesis, you rolled the paper in the typewriter and typed the assignment for the professor to do the evaluation. Word processing has developed from its early days, and processing your writing to the final form has become quicker and easier. At the same time, the sources that are used in higher education studies have changed their form. Previously, writing required students to visit the library, where it was possible to explore scientific publications, books and journals. Today, the world wide libraries have come to student's laptops and tablets. Universities provide library services for their students, and libraries rely more and more on electronic resources. A vast selection of scientific journal articles is available for the student just by a click of a mouse. (Moore 2010.)

Ondrusek (2012) has conducted a literature review about students' writing skills. Even if there is not much research evidence available about how students in higher education master (scientific) writing, she was able to make conclusions about students' writing skills. She reminds us that writing plays an integral part in higher education, *"and fluency with basic writing skills is a pre-requisite to advanced academic writing. This fluency depends on a student's understanding of how to select a topic, organize ideas, and employ rules of writing mechanics"*. Writing skills develop while studying and with support from teachers. (Ondrusek 2012, p. 184.)

In Finland, all higher education institutions provide writing guidelines on their websites and these guidelines are supposed to be used when students write and publish their theses. The guidelines are fairly homogeneous. Students are taught the referencing practices, they are advised to refer to the sources they use and give detailed information about the source in the reference list. In all guidelines and instructions, plagiarism is not accepted in any circumstances. All higher education institutions have agreed to follow national guidelines about good scientific practice. Plagiarism is defined as follows: *"to present someone else's research plan, manuscript, article or text, or parts thereof, as one's own"* (National Advisory Board on Research Ethics, 2002, p. 21).

However clear this definition of plagiarism may sound, to understand it seems to be problematic. Official figures of plagiarism are minimal: The statistics provided by National Advisory Board on Research Integrity (2012) recognised only two cases of plagiarism in Finland in 2011. This is in sharp contrast with international research results. For example Blum (2009) refers to studies conducted in US where 68% of college students admit copying text from the internet. In Finland's neighbouring country Sweden, in 2011 altogether 517 cases of sanctioned student plagiarism cases in higher education were reported by Högskoleverket (2012, p. 7).

There is no research in Finland about how higher education students master academic writing or referencing techniques, neither is there generalizable research on plagiarism. Sloppy scholarship which in plagiarism research and literature refers to writing in which the author fails to give the reader reliable references in text has

neither received any interest in Finland. With no or only a few addressed plagiarism cases annually Finland has become a peculiar country among other countries that publicly try to tackle the problem of plagiarism. In individual suspected plagiarism cases reaching consensus over what is plagiarism and not acceptable is complicated and has led to conflicts within the university (e.g. Moore 2008). These kinds of conflicts have been analysed in many publications about plagiarism in higher education (e.g. Martin 2008; Larkham & Manns 2002; Posner 2007).

Electronic publishing of students' theses has become common in Finland. Universities have their own publication forums, and some universities publish Master's theses electronically on their web pages. The universities of applied sciences (UAS), on the other hand, have a common publication forum, Theseus, where students of all 25 UASes can publish their theses (Theseus 2013). When Kämäräinen (2012) analysed the use of Wikipedia as a source in UAS students' theses he recognised over 30 different types of information sources the students used in their theoretical frameworks. He also made remarks of the reference lists in the theses. They were unfinished to that extent that Kämäräinen (2012) ends up stating: *"It is obvious that the reference lists of the theses in data had not been checked and this can only mean that not even the teachers read them."*

Methodology

The study aims to clarify the referencing practices of Finnish higher education graduates in their theses. The accuracy and consistency of referencing in theses are analysed. The data consist of a convenient sample of theses published in Theseus data base. In early 2013, 40000 publications from 25 different universities of applied sciences are available. For the purpose of this study, altogether 48 theses in the area of health and 43 theses in the area of business studies have been scrutinised in order to determine the quality of referencing. Both Bachelor's theses and Master's theses are included in data. From every institution, the sixth latest Bachelor's thesis in the area of nursing (or related area in one institution) was chosen (n=24) and the third latest Master's thesis in the area of health (n=24). The same procedure was used in the area of business studies resulting to have both Bachelor's (n= 23) and Master's (n=20) theses in the data. The collection of data took place in September–November 2012 (the area of health) and December 2012–January 2013 in the area of business. All the theses were published in 2012.

The formation of the analytic tool was inductive. First, an overlook on the first texts was taken, and notes were made concerning referencing style within text and the formation of lists of references. After examining ten different theses in the area of health, the analysis frame was fixed. From every thesis in the sample the following was checked: referencing practice in text (consistency, accuracy) and the list of references (consistency, accuracy). If inconsistency or inaccuracy was noticed in referencing style or in the referencing list, some plagiarism check-up was taken using Google to identify plagiarism. Copying of pictures and figures was also paid attention to, but the analysis of this is excluded as this would deal with Finnish copyright law and different interpretations of copyright law, which cannot be dealt with in this context.

Table 1

Accuracy and consistency of referencing in students' theses (n = 91)

	Accurate and consistent	Some inconsistency	Constant inconsistency	Failed referencing/Plagiarism
Bachelor of Health Care (n = 24)	16	5	3	0
Master of Health Care (n = 24)	8	2	8	6
Bachelor of Business Administration (n = 23)	18	—	3	2
Master of Business Administration (n = 20)	8	6	3	3
Total number of cases (n = 91)	50	13	17	11
Total (100%)	55%	14%	19%	12%

The main results are presented first in quantitative form, and after that authentic examples of sloppy scholarship and plagiarism are given. The examples have been translated from Finnish to English except in case of word-to-word plagiarism. The names of the students whose theses have been analysed or the names of the institutions that have accepted and published these theses are not mentioned here to avoid whistleblower reactions. In the analysis, comparisons are made to texts that are understood to be original (published earlier). These sources are listed separately after the reference list.

Quality of referencing practice

Referencing practices of altogether 91 theses were analysed. The quality of referencing practice was classified into four categories: **accurate and consistent referencing** (both referencing in text and the reference list were consistent and accurate), **some inconsistency in referencing** (less than ten inconsistent or inaccurate entries noticed), **constant inconsistency in referencing** (ten or more inconsistent or inaccurate entries noticed) and **failed referencing/plagiarism** (at least three paragraphs in text were word to word copies of other author's text and/or the student made false references at least three times). The main result of the analysis is presented in table one.

The results show that the majority of students master referencing practice. In 50 (55%) theses referencing followed the generally accepted norms of academic writing. The reader could rely on the references and find enough information about the publication in the reference list. There were detailed references to the sources that were used and the use of citations was correct. The sources that were referred to in text could be found in the reference lists in alphabetical order and the reference list provided relevant information about the source. Occasional mistakes, e.g. in spelling or in the reference list details were considered acceptable.

In the category of **some inconsistency in referencing** the texts were not completely finalised, but major mistakes were not noticeable. In 13 (14%) theses there was some carelessness in the text like mistakes in alphabetical order or year of publication in reference lists, referrals to web-pages instead of referring to the author, flaws in informing the reader the full information of the reference, “forgetting” to name the page of the cited source, or some references that were mentioned in text might not be found in reference lists.

The number of theses in which the referencing practice was unclear was significant ($n=28$). The two last categories, **constant inconsistency in referencing** and **failed referencing/plagiarism** are analysed further in the following chapters under the titles of sloppy referencing and plagiarism. These two categories are overlapping, and only very clear word to word copying and misleading references were classified as plagiarism. In 17 of the theses (19%) flaws in citing and referencing were constant (the referencing style changed in the text, there was excessive use of secondary/tertiary sources, no details of the source were given so that the reader cannot find it, etc.). Clear-cut plagiarism was identified in 11 theses. Originally, plagiarism was not the main concern of this study and thus plagiarism was identified by putting exact sentences to Google search only after noticing constant inconsistency in references or reference lists. The 11 cases of plagiarism only represent the very obvious cases that were noticed “by plain eye”, and no conclusions about the amount of plagiarism in students’ published theses can be made. It is not known if electronic plagiarism detection had been used before publishing these theses in Theseus.

Sloppy referencing

Sloppy referencing refers to writing where the author fails to present clearly the source s/he has used. There were many kinds of inaccuracies in referencing and I have categorized them into five categories: undetailed citation, confusing reference, quotations without quotation marks, the primary source not identified, and misquoting.

In an **undetailed citation** the student writes about a detailed fact or a research result. Exact information is given without exact reference. The student has a reference in the text, but it is not detailed enough, for example the page number is not given. This type of citation was more common in theses in the area of health. Below is one translated paragraph as an example of undetailed citation practice. There are 51 sources used in this 48 pages long thesis but no page numbers at all are used in referrals to these sources.

During surgery, heating of the patient, fluids and gas placed intra-abdominally is used (Kairaluoma 2007). The use of wetted and heated gases may reduce post-operative pain, and the need for opioids as well as medication for nausea (White et al 2007). Thermal blankets and mattresses are also used (Rotko 2012). The heat balance should be measured during surgery (Kitching & O'Neill 2009).

Confusing referencing is in question when the student gives information about the source in different ways in different parts of the thesis. The following is an example of a thesis in which the student had copied a figure from an internet source. The reference is given, but the number of the page is missing. The title of the figure (translated)

Table 2
Example of direct quotation without quotation marks or reference

Student in 2012, one paragraph:	Ministry of Finance 2008, sentences in different paragraphs:
Asiakkuuden johtamisen tavoitteena on asiakkuuksien arvon nostaminen. Asiakkuuksien johtaminen toteutetaan asiakkuusprosesseja analysoimalla ja luomalla asiakkuusstrategiat, joiden avulla asiakkuuksia voidaan kehittää. Asiakaslähtöisyys edellyttää, että palveluita tarjoava yritys analysoi jatkuvasti asiakaskantaansa ja kehittää vaihtoehtoisia tapoja hoitaa asiakkuuksiaan. Asiakkuutta tulee hoitaa yrityksen asiakkuusstrategian mukaisesti. Asiakkuutta tulee seurata, raportoida ja analysoida säännöllisesti. Prosessin lopputuloksena syntyy selkeä ymmärrys asiakkuuden tavoitteista ja niistä toimenpiteistä, joilla tavoitteet saavutetaan.	"Asiakkuuden johtamisen tavoitteena on asiakkuuksien arvon nostaminen." "Asiakkuuksien johtaminen toteutetaan asiakkuuksia analysoimalla ja luomalla asiakkuusstrategiat, joiden avulla asiakkuuksia voidaan kehittää." "Asiakaslähtöisyys edellyttää, että palveluita tarjoava organisaatio analysoi jatkuvasti asiakaskantaansa sekä kehittää vaihtoehtoisia tapoja hoitaa asiakkuuksiaan." "Asiakkuutta tulee hoitaa asiakkuusstrategian mukaisesti. Sitä tulee seurata, raportoida ja analysoida säännöllisesti. Suunnittelun lopputuloksena syntyy selkeä ymmärrys asiakkuuden tavoitteista ja toimenpiteistä, joilla tavoitteet saavutetaan."

is: “Figure 4. Sources of every day customer information behind experiencing value (Korkman & Arantola 2009)”. In the list of references the student writes the names of authors in different order: “Arantola, H.& Korkman, O.” The reference is accurate, but there is inconsistency between the reference in text and the reference list. In his text the student has referred to different pages of Arantola & Korkman 2009, to Korkman & Arantola without a year or page numbers and to Korkman et al. 2009 to a page. The Google search was confusing as well. There was another thesis in which, quite in the same way, references were made to Korkman & Arantola, but in the reference list there was Arantola & Korkman. Later in the Google search it was found out that the source Arantola & Korkman does not exist, and the right reference should consistently be made to: Korkman, O. & Arantola, H. 2009.

In the data there are numerous instances where the students write **quotations without quotation marks**. If any kind of reference was given this was seen as sloppy referencing, but in case of missing references this kind of writing practice was considered plagiarism. In table 2 one example is presented. The copying is almost word to word and thus the example has not been translated. Exactly same wording is highlighted in student’s text and the missing highlighting shows where a synonym has been used.

Similar sentences were found in a document published by the Ministry of Finance presenting the customer approach of national IT service unit. The student has picked sentences from the original document and formed paragraphs consisting of copied sentences. There are no references at all in the student’s text when she writes about customer management (two pages) and thus this is an example of plagiarism as well.

In the data it is common that **the primary source is not identified**. Thus secondary sources and even tertiary sources are used also in central parts of the

Table 3

Example of thesis text in which primary source is not identified

Student	VirtuaaliAMK
Hyvinvointi käsitteen määrittäminen on vaikeaa, koska erilaiset asiat tuottavat eri ihmisille hyvinvointia. Se ymmärretään helposti vain aineelliseksi elintasoksi, vaikka kysymys on laajemmasta kokonaisuudesta. Hyvinvoinnissa nähdään yksilöllinen, yhteisöllinen ja yhteiskunnallinen ulottuvuus. Se on myös kulttuurisidonnainen asia; opitut ja omaksutut yhteiset tavat ja normit, käyttäytymissäännöt, arvot, asenteet ja arvostukset heijastuvat ihmisten hyvinvointikäyttäytymisessä. Ihmisten hyvinvointi on ylin tavoite, johon koko yhteiskuntapolitiikka ja erilaiset palvelut pyrkivät. (Virtuaalinen AMK 2012.)	"Hyvinvointi-käsitteen määrittelemineen on vaikeaa, koska erilaiset asiat tuottavat eri ihmisille hyvinvointia. Se ymmärretään helposti vain aineelliseksi elintasoksi, vaikka kysymys voi olla laajemmasta ilmiöstä. Hyvinvoinnissa nähdään yksilöllinen, yhteisöllinen ja yhteiskunnallinen ulottuvuus. Se on myös kulttuurisidonnainen asia: opitut ja omaksutut yhteiset tavat ja normit, käyttäytymissäännöt, arvot, asenteet ja arvostukset heijastuvat ihmisten hyvinvointikäyttäytymisessä. "Ihmisten hyvinvointi on ylin tavoite, johon koko yhteiskuntapolitiikka ja erilaiset palvelut viime sijassa pyrkivät."

theses. Accordingly, referencing remains unclear. It is also common to use general introductory internet pages as a source. Table 3 presents one example where the student writes about the main concept of her thesis: welfare. The text is a copy of study material in VirtuaaliAMK (Virtual University of Applied Sciences). She has quoted text without quotation marks. The study material has some references but the student ignores them, and she also makes a mistake in naming the source. The copying here is (almost) word to word with minor changes to the text.

Misquoting means that the reference given in text is incorrect and misleading. At one page in her 50 page Master's thesis the student refers to an article by Almeida et al. 2001 in the following way (text translated):

Irritability, fatigue and stress may prevent the parent to notice the child's needs or tighten the relationship between them. Even good educational principles do not necessarily help, if the parent loses his/her head due to stress and exhaustion. (Almeida, Wethington and Macdonald 2001.)

There is no other reference in the thesis to this source. In the reference list the student writes:

Almeida, D., Wethington, E. & Macdonald, D. 2001. Daily variation in parental engagement and negative mood: implications for emotionally supportive and conflictual interactions. *Journal of marriage and the family* (63), 471–460.

The first thing the reader starts to wonder are the page numbers as they are given backwards. The correct publication details are: *Journal of Marriage and Family* 63 (2), 417–429. In this short paragraph the student has made many mistakes. *Paternal* in the title has become *parental*, there are mistakes in the title of the journal and in reporting the volume and number of the issue. The page numbers are incorrect. Additionally, the content the student has written is not from this article. The origin of the text remains unknown.

Another example of misquoting or a false reference (text translated) is presented below:

The minimum number of nursing staff is significant above all for nursing staff stability in the organisation. It has been shown that for each additional patient i.e. increase in workload effects nursing staff 15 per cent increase in job dissatisfaction and 23 per cent increase in probability of burnout. Thus, the use of substitutes has significance in the skilled personnel for work and to coping at work. (Pitkääho 2011, 113; Aiken, Sloane, Cimiotti, Clarke, Flynn, Seago, Spelt & Smith 2010, 904–906.)

The student has referred to two original texts, Pitkääho's (2011) doctoral dissertation (in Finnish) and to an article by Aiken et al. 2010. Neither of the references is correct. On page 113, Pitkääho (2011) does not write about the issue the student writes in the paragraph, and in her dissertation there is no concept of 'minimum number of staff' or 'job dissatisfaction'.

In the student's 67 page long thesis there is only one reference to Aiken et. al. 2010. She refers to pages 904–906 of the article. However, in the Aiken et al. (2010) article those first pages deal with nurse ratios in three USA states, and the relation between nurse-patient ratio and patient mortality. Thus, the issues the student writes about in her text cannot be found on those pages (nor in the whole article). It remains unclear where the student has taken the exact percentages to her paragraph and why references are made in a misleading way to those two sources that do not contain the information.

Plagiarism

Several plagiarised text columns were also found within the sample ($n = 91$). In altogether 11 theses plagiarism was identified without using plagiarism detection software. Plagiarism was evident and typically it was cut and paste plagiarism, text was copied from several sources. In many cases some of the wording had been changed: synonyms were used, the sentences were shortened or word and sentence order changed. In table 4 one example of plagiarism is presented.

The text and the references have been copied word to word. The student has also copied the references from the original authors' reference list. A further proof of copying is that also the mistakes that are understandable in a given text have been transferred to the copied text ("researcherer"). The copied link in student's reference list, however, does not work due to removing -. The link in the original authors' reference list is correct.

Another example of plagiarism represents recycling of text published earlier in the same data base Theseus (Table 5). In a similar way as in the previous example, the reference is copied word to word and the reference list is identical except a spelling mistake made in the word "Publisherss".

Conclusion

Students learn to write and cite in their higher education studies, and they become familiar with at least one referencing style that the university or study field in question prefers to use. The analysis of a sample of theses ($n=91$) published in the Finnish Theseus data base produced knowledge about referencing skills of graduating students in universities of applied sciences. The majority of graduates in this sample (55%) master the recommended referencing procedure and in that way are able to follow an

Table 4

Example of plagiarism in theses

A Master's thesis in health care, one author, 2012	A Master's thesis in early childhood education (Korppi & Latvala 2010)
<p>Jones (2004, 129) muistuttaa, että sellaiset rakenteet, jotka mahdollisesti rajoittavat lasten osallistumista, on muutettava sellaisiksi, että osallistuminen on mahdollista ja merkityksellistä. On myös tärkeää hahmottaa YK:n lapsen oikeuksien sopimuksen artiklat lapsen oikeutena osallisuuteen eikä niinkään pakkona. Lapselta ei näin ollen viedä lapsuutta ja anneta hänelle liikaa vastuuta päätöksentekotilanteissa, vaan hänelle annetaan oikeus osallisuuteensa ikään ja kehitystasoonsa nähden. (Lansdown 2001, 8.)</p> <p>In reference list: Jones, A. 2004. Involving children and young people as researchers. Teoksessa S. Fraser, V. Lewis, S. Ding, M. Kellet & C. Robinson. (toim.) Doing research with children and young people. Lontoo: SAGE, 113–130.</p> <p>Lansdown, G. 2001. Promoting children's participation in democratic decision-making. New York: Unicef. PDF-dokumentti. http://www.unicef.org/publications/pdf/insight6.pdf Luettu 14.12.2011</p>	<p>"Jones (2004, 129) muistuttaa, että sellaiset rakenteet, jotka mahdollisesti rajoittavat lasten osallistumista, on muutettava sellaisiksi, että osallistuminen on mahdollista ja merkityksellistä. On myös tärkeää hahmottaa YK:n lapsen oikeuksien sopimuksen artiklat lapsen oikeutena osallisuuteen eikä niinkään pakkona. Lapselta ei näin ollen viedä lapsuutta ja anneta hänelle liikaa vastuuta päätöksentekotilanteissa, vaan hänelle annetaan oikeus osallisuuteensa ikään ja kehitystasoonsa nähden. (Lansdown 2001, 8.)"</p> <p>In reference list: Jones, A. 2004. Involving children and young people as researchers. Teoksessa S. Fraser, V. Lewis, S. Ding, M. Kellet & C. Robinson. (toim.) Doing research with children and young people. Lontoo: SAGE, 113–130.</p> <p>Lansdown, G. 2001. Promoting children's participation in democratic decision-making. New York: Unicef. Luettu 13.01.2010 http://www.unicef-irc.org/publications/pdf/insight6.pdf</p>

ethically sound writing practice. The theses are written in a way that they show student learning about their chosen subject and are proofs of student expertise. However, in this sample almost one third (31%) of the graduates did not follow good writing and referencing practice in their theses. This result poses questions to universities of applied sciences: What kind of learning do these theses show and what kind of text is accepted as a thesis in higher education?

There is not much discussion about plagiarism in Finnish higher education, and officially, the only authority reporting plagiarism is Finnish Advisory Board on Research Integrity. In the context of higher education discussion about plagiarism has proven to be difficult. In the data presented in this paper there is evidence of student plagiarism that is accepted in universities of applied sciences. The results show how text from existing source is taken and used as student's own writing. Sentences might or might not be slightly altered, sources of the original text are copied to reference lists, and nothing convinces the reader that the student herself/himself has written the text based on learning due to reading the named source. Also, many of the practices categorised under the blurry concept of sloppy scholarship would be seen as plagiarism in many other countries. In theses that include plagiarism it is difficult to find learning in which students have constructed knowledge for themselves (Barr & Tagg 1995, p. 15) instead of just copying text from one document to another.

Table 5

Example of plagiarism: recycling text within *Theseus*

A Master's thesis in 2012, one author	A Bachelor's thesis, (Partanen 2010)
Osaamisen johtaminen on tullut laajemmin esille viimeisten 8-10 vuoden aikana. Voisi melkein pä sanoa, että osaamisen johtamisesta on alettu keskustelemaan asteittain, ja siten tuotu ihmisten keskuuteen. Osaamisen johtaminen on noussut kuumaksi puheen aiheeksi yritysten keskuudessa. Se on näkyvästi verkottunut myös talouteen. Osaamisen johtamista voidaan tarkastella myös lisääntyneen kilpailun, fuusioiden ja uusien hankintojen myötä. (Gamble, Blackwell 2001, 6-7.)	"Osaamisen johtaminen on tullut laajemmin esille viimeisten 8-10 vuoden aikana. Voisi melkein pä sanoa, että osaamisen johtamisesta on alettu keskustelemaan asteittain, ja siten tuotu ihmisten keskuuteen. Osaamisen johtaminen on noussut kuumaksi puheen aiheeksi yritysten keskuudessa. Se on näkyvästi verkottunut myös talouteen. Osaamisen johtamista voidaan tarkastella myös lisääntyneen kilpailun, fuusioiden ja uusien hankintojen myötä. (Gamble, Blackwell 2001, 6-7.)"
In reference list: Gamble, P. & Blackwell, J. 2001. Knowledge management: a state of the art guide, Kogan Page Publishers. http://books.google.fi/books?id=1Dj09GVXOVgC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false	In reference list: Gamble, P. & Blackwell, J. 2001. Knowledge management: a state of the art guide, Kogan Page Publishers. http://books.google.fi/books?id=1Dj09GVXOVgC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

This study revealed plagiarism and sloppy scholarship in theses that have been accepted as final assignments for Bachelor's or Master's degrees. Plagiarism in thesis texts was identified by using the definition of Finnish Advisory Board on Research Ethics (2002, 21) when "*someone else's text, or parts thereof*" were "*presented as one's own*". As there is neither research about the amount of plagiarism nor practically any officially recognised plagiarism cases in Finland, the evidence produced in this study needs careful consideration. The conclusions also need to be careful as there is proof of strong reactions with ad hominem accusations towards the rare whistle-blowers in Finland (Moore 2008). The evidence is in sharp contrast with the educational discourse in Finland in which quality is the key concept.

The results, however, suggest that Finnish higher education disagrees with On-drusek's (2012) conclusion that a student with deficiencies in writing skills is not able to fulfill degree requirements, and that there might be a hidden problem of plagiarism in Finnish higher education. The results can also mean that in Finnish higher education plagiarism is understood and defined differently than in the national definition or in international literature. There seems to be variation in how plagiarism is understood and in the criteria for an acceptable writing practice.

References

- Apastyle.org. (2013) *APA Style Publication Manual*. American Psychological Association
<http://www.apastyle.org/> [accessed 16 February 2013]
- Barr, RB. & Tagg, J. (1995) From Teaching to Learning—A New Paradigm for Undergraduate Education. *Change* 27 (6) 13–25.
<http://docushare3.dcc.edu/docushare/dsweb/Get/Version-3000/BarrTagg.pdf> [accessed 16 February 2013]

- Blum, SD. (2009) *My Word! Plagiarism and College Culture*. Cornell University Press
- Finnish Advisory Board on Research Ethics (2002). *Good scientific practice and procedures for handling misconduct and fraud in science*.
http://www.tenk.fi/sites/tenk.fi/files/Hyva_Tieteellinen_ENG.pdf [accessed 16 February 2013]
- Finnish Advisory Board on Research Integrity (2012) *Annual report 2011*.
http://www.tenk.fi/sites/tenk.fi/files/tenk_toimintakertomus_2011.pdf [accessed 19 April 2013]
- Högsköleverket (2012) *Disciplinärenden 2011 vid högskolor och universitet*. Rapport 2012:19 R <http://www.hsv.se/download/18.485f1ec213870b672a680004302/1219R-disciplinarenden-2011.pdf> [accessed 21 April 2013]
- Larkham, P.J. & Manns, S. (2002) Plagiarism and its treatment in Higher Education. *Journal of Further and Higher Education* 26 (4), 339–349
- Martin, B. (2008) Plagiarism Struggles. *Plagiary: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification*. <http://www.bmartin.cc/pubs/08plagiary.html> [accessed 21 April 2013]
- Masters, K. (2005) Flawed evidence: a case study of misquoting and inaccurate referencing. *South African Journal of Libraries & Information Science* 71 (3), 282–288
- Moore, E. (2008) The Four Stages of Addressing Plagiarism. *3rd International Plagiarism Conference*.
<http://www.plagiarismadvice.org/media/2008papers/P34%20Moore.pdf> [accessed 19 April 2013].
- Moore, E. (2010) Plagiointia Suomen korkeakouluissa? (Plagiarism in Finnish higher education?)
Kever-Osaaja 1(2). <http://www.uasjournal.fi/index.php/K-O/article/viewFile/1230/1142> [accessed 19 April 2013]
- Ondrusek, AL. (2012). What the Research Reveals about Graduate Students' Writing Skills: A Literature Review. *Journal of Education for Library & Information Science* 53 (3), 176–188
- Park, C. (2003) In Other (People's) Words: plagiarism by university students—literature and lessons. *Assessment & Evaluation in Higher Education* 28 (5), 471–488
- Posner, RA. (2007) *The Little Book of Plagiarism*. New York: Pantheon Books
- Sutherland-Smith, W. (2008) *PLAGIARISM, The Internet and student learning. Improving Academic Integrity*. New York: Routledge
- University of Cardiff (2013) *Citing and Referencing in the Harvard Style*.
<http://www.cardiff.ac.uk/insrv/resources/guides/inf057.pdf> [accessed 16 February 2013]
- University of York (2013) *MLA referencing Style*. <http://www.york.ac.uk/integrity/mla.html> [accessed 16 February 2013]

Sources used in text comparisons

- Aiken, LH., Sloane, DM., Cimioti, JP., Clarke, SP., Flynn, L., Seago, JA., Speltz, J. & Smith, HL. (2010) Implications of the California nurse staffing mandate for other states. *Health services research* 45 (4), 904–921.
- Almeida, DM.; Wethington, E. & McDonald, DA. (2001) Daily Variation in Paternal Engagement and Negative Mood: Implications for Emotionally Supportive and Conflictual Interactions. *Journal of Marriage and Family* 63 (2), 417–429
- Korppi, M. & Latvala, J. (2010) *Osallistava vuorovaikutuksen malli lasten osallisuuden mahdollistajana*. Master's thesis in early childhood education. University of Jyväskylä. <https://jyx.jyu.fi/dspace/bitstream/handle/123456789/25307/URN%3ANBN%3Afi%3Aju-201009302820.pdf?sequence=1> [accessed 16 February 2013]

- Ministry of Finance (2008) *Valtion IT-palvelukeskus ASIAKKUUSUUNNITELMA 2008*.
www.vm.fi/.../20080707Valtio/VIP_asiakkuussuunnitelma_v2.1.doc [accessed 16 February 2013]
- Partanen. MR. (2010) *Osaamisen johtaminen Helsingin OP Pankissa*. Bachelor of Business Management Thesis, Laurea UAS. http://publications.theseus.fi/bitstream/handle/10024/24697/Partanen_Milla.pdf?sequence=1 [accessed 19 April 2013]
- Pitkäaho, T. (2011) *Hoitotyön henkilöstömitoitus ja tulos kompleksisessa erikois-sairaanhoidon toimintaympäristössä*. University of Eastern Finland. http://epublications.uef.fi/pub/urn_isbn_978-952-61-0395-2/urn_isbn_978-952-61-0395-2.pdf [accessed 19 April 2013]
- VirtuaaliAMK (2012) *Terveysten ja hyvinvoinnin edistämisen tietoperusta (Ylempi AMK)*. Hyvinvointi. <http://www2.amk.fi/digma.fi/www.amk.fi/opintojaksot/030708/1080212137262/4WJJVzINI/4WJJVOXoe/4WJJVXbh8.html> [accessed 19 April 2013]

Author

Erja Moore, erja.moore@karelia.fi, Karelia University of Applied Sciences, Finland

Copyright © 2013 Author Erja Moore: The author grants to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

STUDENTS' EXPERIENCE OF INSTITUTIONAL INTERVENTIONS ON PLAGIARISM: NIGERIAN CASE

Stella-Maris Orim, Erik Borg, Isabella Awala-Ale

Abstract: In the last ten years, there has been much research into academic integrity with a focus on plagiarism in developed countries. There is still a dearth of such research in developing countries like Nigeria.

This paper presents the results from a larger exploratory study on student plagiarism in Nigerian Higher Education Institution of Learning (Nigerian universities), which is associated with the Impact of Plagiarism Policies in Higher Education Across Europe (IPPHEAE) project.

It seeks to contribute to the body of knowledge by focusing on the experiences of previous institutional interventions on student plagiarism issues and their impact on student experience when they study elsewhere.

Nigerian students studying abroad had to adapt to diverse teaching, learning and assessment styles under a different institutional system. This resulted in the students struggling when they had to apply skills they had not acquired during their previous study.

This study adopted a mixed method approach; 25 Nigerian Postgraduate Students studying in a United Kingdom University were interviewed for the qualitative data and 171 IPPHEAE student questionnaires were completed for the quantitative data.

Results from the data suggest that the previous institutional system experienced by the students was quite different from what they met in their present institutions of study in England. As a result, the students struggled to cope with their studies when they had to apply skills they had not acquired.

Key words: Nigerian universities, IPPHEAE, students' experience, student plagiarism, student mobility

Introduction

Currently, due to the industry requirements in Nigerian organizations and a drive for professional improvement, there has been an increase in the influx of Nigerian students to overseas universities for postgraduate studies. These students were classified as international students, relative to their country of origin. Deckert (1993) referred to international students as a group of students for whom English is not their language of origin and who regularly risk being accused of plagiarism. Because of the shared desire for success on the part of the students and the institutions in which they have been accepted, there is a need to understand their previous academic background, the challenges they face while studying in the new academic background and the impact the previous experience has on their overall learning experience.

This study seeks to add to the recent study in this area by exploring the previous experiences of Nigerian postgraduate students in a United Kingdom university. This research aims at contributing to knowledge in the area of the impact of previous institutional interventions on the academic experience of Nigerian students who travel overseas to further their education.

Literature Review

One of the principles for effective teaching and learning is the promotion of the active engagement of the student as learner. Active engagement takes into account the various types of learning styles of the students and ensures that each is catered for in the course of study. David (2012: 18) suggests that the main aim of higher learning should be learners' independence and autonomy. This, she believes, is evidenced in the "active engagement of students in their learning, ensuring they acquire a repertoire of learning strategies and practices, develop positive learning dispositions and build the confidence to take ownership of their learning". When this is achieved, deep learning is said to have taken place. In line with this, social development theory suggests that when learning takes place, there is the ability to make an inner form of transfer from what is learnt to what is said or written for assessment (Vygotsky 1978). This theory suggests that the learning from an assessment can be compromised by the students in several ways, which will result in the transfer failing to take place or being ineffective. There are several forms of academic misconduct that result to the compromise of assessment results. Some of these are: cheating, collusion and plagiarism. This study has focused on plagiarism.

Park (2003) framed plagiarism as a form of academic malpractice. He describes it as a breach of academic integrity and defines it as "the theft or words or ideas, beyond what would normally be regarded as general knowledge" (Park, 2003: 472).

Like Park (2003), Carroll defines plagiarism as "submitting someone else's work as your own" (2007: 13). She further said that despite the fact that she defined it so simply; the actual phrase may represent more meaning than it appears to. She explained "submitting" in terms of handing in some work for the purpose of academic credit. "Someone else's work" was depicted as inappropriately referenced piece for submission. So many attempts have been made at the definition and other studies have shown that different understandings of the same word are held by teachers and students (Delvin 2003; Shi 2004). This raises a need for consistency in its definition and application.

Studies carried out in this area further suggest that most students who admit plagiarising suggest that it occurred as a result of issues including: "... lack of intent"; "... lack of awareness"; "... lack of the proper skills". These instrumental conceptualizations contrast with viewing plagiarism as an ethical issue, which places it in the arena of ethical reasoning. It is well known that the "ethical climate" (Victor and Cullen, 1993) of an organization represents the shared perceptions of what ethically correct behavior is and how ethical issues should be handled.

Organizational ethical context is basically the factors affecting the ethical judgment of people in an organisation (Pasternak, nd). The two main areas are the ethical climate and the ethical culture. Organizational ethical culture is the way things are done in a given organization and represents a whole range of several "formal" and "informal" systems of behaviour control capable of resulting in either ethical or unethical behavior (Trevino et. al, 1998).

Some of these formal systems are factors such as training programs, policies, authority structure, leadership, reward systems, penalty systems, while the informal systems include ethical standards, peer behavior and influence from teachers and students.

The ethical decision-making literature considers ethical culture to be a significant component in decision-making processes (Trevino, 1986; Hunt and Vitell, 1986) which include the directions for day-to-day behavior (Cohen, 1993) helping establish what is considered to be legitimate or unacceptable in an organization or institution (Trevino and Ball, 1992).

The ethical climate as opposed to the ethical culture is viewed as the shared perceptions of what behavior is acceptable and of how ethical issues should be handled in the organization (Victor and Cullen, 1993). The ethical climate is described by Kohlberg (1969) as having three levels of moral judgment and sphere of analysis, which can be individual, local or international. According to Victor and Cullen (1993), the ethical climate reflects the criteria individuals use in ethical decision making, which in this context is “acknowledging the use of other people’s words and ideas”. Given that the students studied in institutional climates in which the use of other people’s words or ideas was not identified as a significant ethical issue, it becomes a great challenge to adapt when they move on to other institutions where the ethical climate was different and which had various ways of preventing, detecting, mitigating and penalizing instances of plagiarism.

The Nigeria National Policy on Education (2004: Section 1 Sub-Section 5) states that Nigeria’s philosophy of education is based on the development of the individual into a sound and effective citizen. It further emphasises the full integration of the individual into the community, with the primary goal of producing a well-rounded individual. However, there have been regularly occurring comments on the lack of implementation of this policy, resulting to a negative impact on the standard of tertiary education in Nigeria.

Recently, there has been a growing awareness of plagiarism and its negative consequences in Nigeria evidenced through incidents that have increasingly exposed plagiarism as a form of behaviour that should no longer be allowed to thrive in the Nigerian educational institutions. These public comments are an attempt to re-shape the “*organizational culture*” of the universities from the outside.

Describing the level of educational awareness of Nigerian students, Arenyeka (2012) stated that a second year student in a British university can write an essay which a fourth year student in a Nigerian university will download believing that since it was placed on the internet, it is of an academically acceptable standard. He went on to state that as lecturers they have issues with students plagiarizing, as it is theft of other people’s original work and such theft is not acceptable. This portrays his views about the level of achievement of students studying in the Nigerian university as opposed to one studying in the British university.

On the aspect of penalizing plagiarism offenders, Nnabugwu (2012) reported that the Independent Corrupt Practices and Other Related Offences Commission (ICPC), in conjunction with the National Universities Commission (NUC) are set to focus on academic related corrupt practices such as collective plagiarism by students and staff of Nigerian universities. There have been several cases of such including the recent dismissal of four lecturers on grounds of plagiarism at the University of Calabar (Kalu, 2013).

Cronbach’s Alpha	N of Items	Cronbach’s Alpha	N of Items	Cronbach’s Alpha	N of Items
.773	13	.773	9	.836	22

Figure 1. Reliability Statistics

Another case was the instance as reported by Chiedozie (2012) who stated that a United States-based Nigerian lecturer has sued the Governor of the Central Bank of Nigeria, Dr. Lamido Sanusi, for allegedly plagiarizing his works. There are several such reported cases and accusations which seemingly depict the negative consequences of institutional systems that may not be properly addressing the issue of plagiarism and appropriate acknowledgement of sources at the fundamental level of learning.

On the way the views of the Nigerian society are changing towards plagiarism, NEWS of the PEOPLE (2013) narrated that the Government of Ogun State was accused of lacking morality because the reputation of the invited speaker (former Irish Prime Minister Bertie Ahern) who was accused of plagiarism in the past, was not considered.

These cases suggest that acts of plagiarism in Nigeria can no longer be swept under the carpet as was the norm in the past; however, the universities still remain the foundation for the inculcation of exemplary conduct in students for relevant research addressing the needs of the society.

Method

For this study, a mixed method approach was adopted; Nigerian postgraduate students studying in the Engineering faculty of a United Kingdom University participated in the data collection process. The qualitative approach involved the collection of data from 25 semi-structured interview sessions, which were analyzed with Atlas.ti (Ver. 6.2). The quantitative data was collected from 171 participants who completed the IPPHEAE student questionnaires. The study was an aspect of a larger one carried out on student plagiarism in Nigerian universities. As a result of the affiliation of that study to the IPPHEAE project, the IPPHEAE student questionnaire was used for the data collection and the aspect that relates to this study was analyzed for this paper.

in the questionnaire, there were two sets of questions for the “institutional interventions in place” and the “plagiarism policy, guidelines and procedures”. These had 9 and 13 sub-questions respectively. Results of the test for reliability: *Cronbach’s* alphas for the 9 questions on “institutional interventions in place” and 13 questions on “plagiarism policy, guidelines and procedures” were .77 and .77 respectively (figure 1). A combination of both questions was found to be highly *reliable* with an alpha of 0.836. (22 items; $\alpha = .836$).

Findings

The results were mixed and comprised both qualitative and quantitative findings. The questions were about students’ experience in their previous academic background; the difficulties they faced with “academic writing”; their views on what procedures their

previous institutions had in place to combat student plagiarism and their views on the occurrence of student self-declared plagiarism (PlagOccur). The data was analyzed to check the correlation between the occurrence of plagiarism and both the “Institutional policies and procedures (InstiPlagPolPro)” and “institutional intervention in place (InstitIntervention)”.

Experience on the previous academic background

From the interview sessions with the participants, it was clear that their previous academic background was quite different from that which they met in the UK university. Practices related to plagiarism were significantly different; simply put, the institutional *Ethical context (climate and culture)* were different in a number of ways. Some academic practices which were acceptable in the Nigerian universities in which they had studied were not acceptable in the UK universities in which they now study. There were also differences in the issue of penalties for plagiarism.

From their comments, it can be seen that while they studied in their Nigerian universities most of the students neither knew nor were concerned about plagiarism. Either because they were not taught, it was not detected in their writing or that they were not penalized. Furthermore, they seemed to think that plagiarism was all about referencing. This is reflected in the comment below:

“During my undergraduate days I never really bothered myself with the issue of plagiarism. In-text citation and referencing of authors was most rarely practised and whenever I cited or referenced, I did so without following any referencing standards (like CU-Harvard, Harvard, Oxford etc)... I did not even know these standards existed until I began my program at AAA Uni...”

A broad overview of their previous background was summarized under the themes: teaching, learning, assessment, academic writing, research, study and ICT skills (figure 2).

Difficulties faced with “Academic Writing”

Results from the interviews suggest that Nigerian post-graduates’ previous experience with writing in an academic manner is not in line with western conventions. Data from the sessions with the students, suggests that a number of them were not taught about the concept of plagiarism in the Nigerian universities they came from. One of the participants referred to his experience in the UK university he now studied as a “progressive journey backwards”, explaining that these are the concepts he should have learnt earlier and build up over time. He expressed himself as follows:

“... professional academic writing till date has been a “progressive journey backwards”. I use these words to summarise my experience as the experience started only when I was writing my undergraduate dissertation back in Nigeria and then continued while studying for my masters’ degree in AAA University.”
“... Its a journey backwards because it’s an experience that takes me back to the fundamentals of academic writing which I would have learned much earlier in my educational background. It is an issue today because it wasn’t inculcated as a culture early enough.”

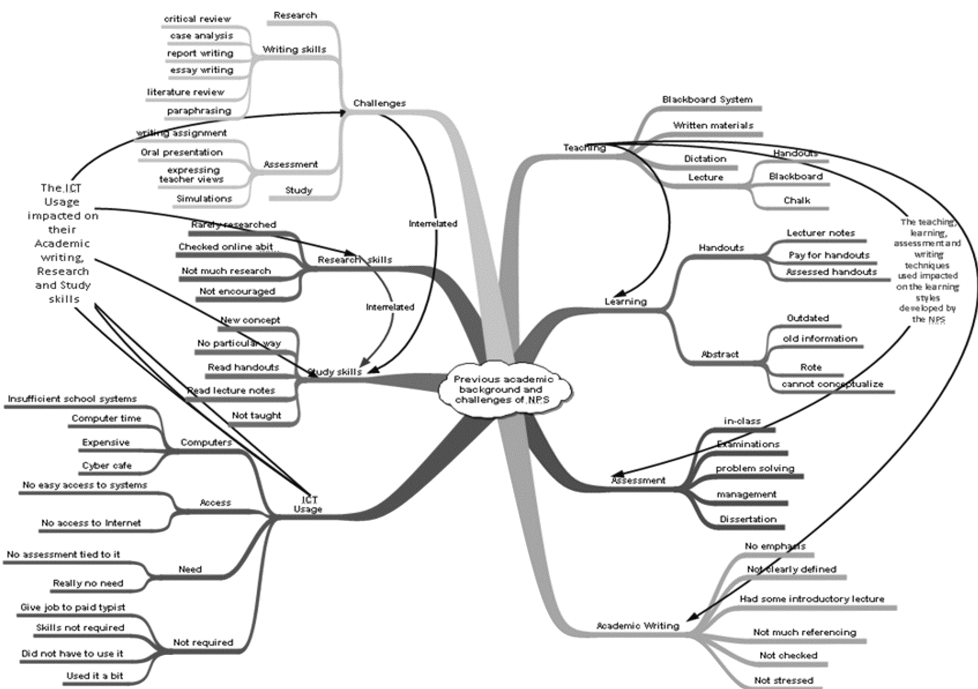


Figure 2. Previous academic background of Nigerian Postgraduate Students

In line with these qualitative findings, the questionnaire investigated what Academic Writing Difficulties (AWD) the participants had. It listed skills such as: 1) Finding good quality sources (AWD_1); 2) Referencing and citation (AWD_2); 3) Paraphrasing (AWD_3) and 4) Understanding different referencing formats and styles (AWD_4).

From the 171 participants who filled out the questionnaires, the result suggests that most of the participants had difficulties with all four skills. From Figure 3, the percentage that admitted to having difficulties with the four skills is higher than half of the sample (67%; 56%; 55% and 58% respectively). This is in line with the qualitative findings.

What their institutions had in place to combat student plagiarism

The results from the survey on the students’ views on NHE Institutional intervention represent answers to questions on the presence of institutional intervention or penalties and the ratings are from agreeing strongly that there are no interventions in place to mitigate student plagiarism, to agreeing strongly otherwise. The penalties in question ranged from verbal warning to expulsion from the institution of study. From figure 4 it is shown that most students felt that their previous institutions in Nigeria had few of these penalties. This is seen in the “over 50%” disagree rate for 9 out of the 13 options (69%).

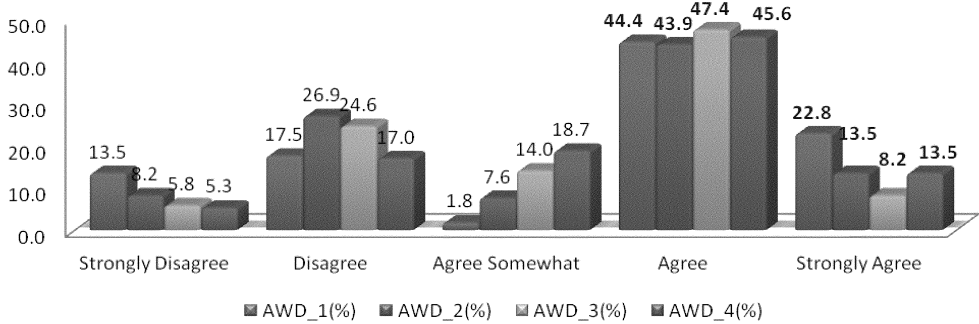


Figure 3. Academic Writing Difficulties (AWD) encountered by NPS

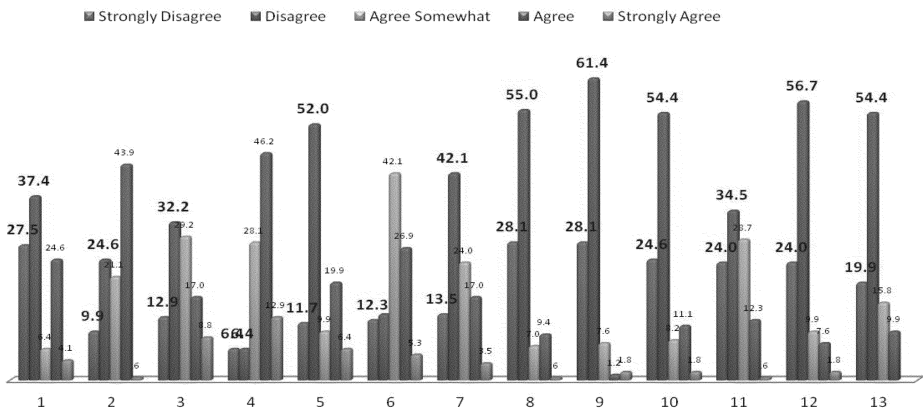


Figure 4. Student views on NHE Institutional intervention

From figure 5, 66% of the students said that they were not aware of any policies in the Nigerian universities they studied. The result also shows that though most of them said they were unaware of these policies and guidelines (figure 5), they still suggest that some form of action will be taken as 64% disagree with the views that “no action will be taken” (figure 6). This raises the question of “what action should be taken?” and “at what level should this action be taken?”

The impact of the institutional plagiarism policies, procedures and guideline on student declared plagiarism

Based on the responses of the students to their previous engagement in plagiarism and the data on the presence of policies, procedures and types of penalties in place, there was a need to know if what is in place is effective.

The impact of the “Institutional Plagiarism Policy and Procedure” (measured by *InstiPlagPolPro*) and “Institution Intervention” (measured by *InstitIntervention*) on

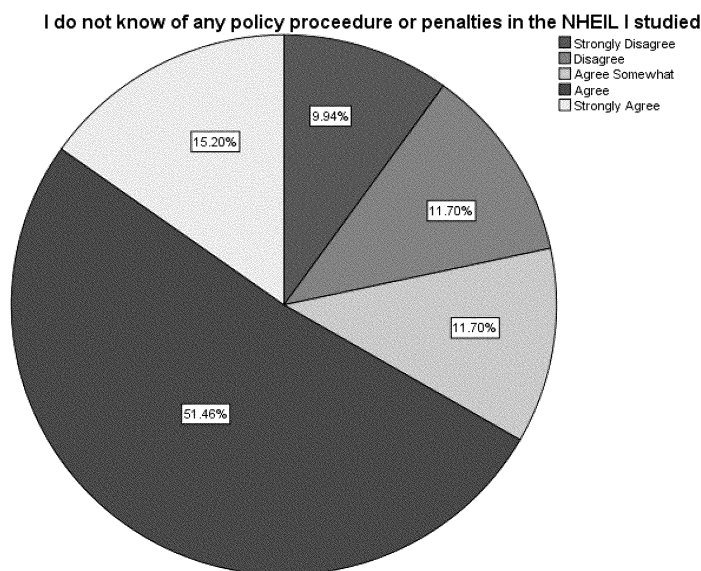


Figure 5. Student views on Nigerian universities policies, procedures and guidelines in place

“occurrence of student self-declared plagiarism” (as measured by *PlagOccur*) was investigated using Spearman’s rho correlation coefficient.

The results for each of the spearman’s correlation analysis are presented in figure 7 and 8.

From the findings, there was little correlation between the two variables and the occurrence of plagiarism. This implied little or no impact of the formal systems in place in the Nigerian universities.

Discussion

A look at a typical Nigerian Higher institution from which these Nigerian postgraduate students came reveals that the pedagogy is quite different from that of the UK universities. In the UK, it is expected that at higher institution level students will be able to take ownership of their learning, while it is apparent from the data that in a number of the Nigerian universities, an approach based on the “conduit” model (Reddy, 1979) of teaching and learning is still commonplace.

As can be seen in Figure 2, the student’s previous academic experience created learning gaps which they had to fill to be able to adapt to studying in UK. Students had issues with basic study skills such as note taking, typing, use of ICT, VLEs, summarizing, paraphrasing, research, referencing styles and sourcing for academic journals.

The data suggests further that the mode of assessment in their Nigerian universities was largely by exams (80%) and there was no real need for substantial development of academic writing until students arrive at the end of their study, at which point, there is a

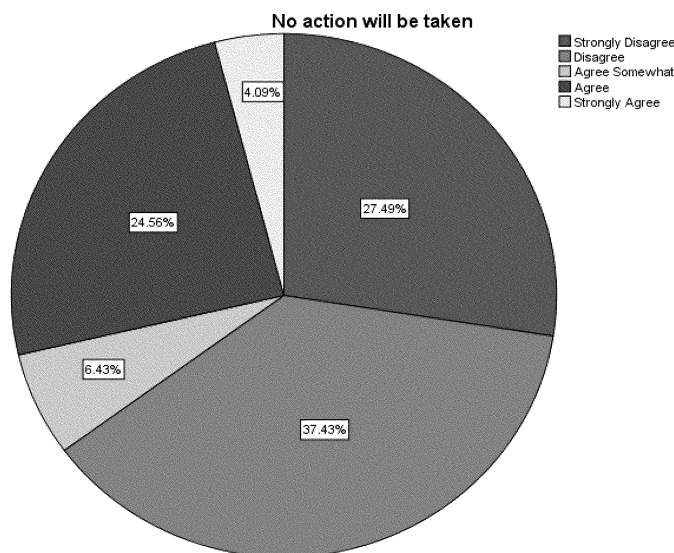


Figure 6. Student views on penalties

dilemma. The students are faced with the probability of getting supervised by lecturers who would not go the extra mile of checking appropriate writing styles or lecturers who have studied outside the country and would check that the written piece of work is up to an acceptable standard.

A few comments from the students suggested that they are not aware of institutional policies in place or being enforced to ensure that the students get fair and consistent assessment in every course in the universities they studied in Nigeria.

When the dissertation is submitted at the end of their study, they are turned in via hardcopies in most cases. With the student-teacher ratios worsening in virtually all disciplines (Oni, 1996, quoting Yesufu, 1996), it becomes a challenge for a teacher to consistently go through each submission in detail and detect instances of plagiarism as there are no digital means of detecting instances of plagiarism.

On arrival, UK universities expect students to start studying immediately with total commitment while some students missed the induction due to the lateness in securing visas and found that they had no options but to learn the basics on their own. The initial shock to most of them was the seeming expectation they perceived from the UK University in which they were studying that every Masters student can make sufficient use of computers for the purpose of studying. During the course of attending the first module/course, they are faced with coursework deadlines. These pieces of coursework were required to be keyed into the system and submitted on-line through Turnitin, the on-line system that searches assignments for similarities to other texts.

Many of these procedures are seemingly new: having to key in the piece of coursework on their own as opposed to “paying typists’ in a business center” in Nigeria; “online soft copy submission . . .” as opposed to “hard copy submissions” that are hard

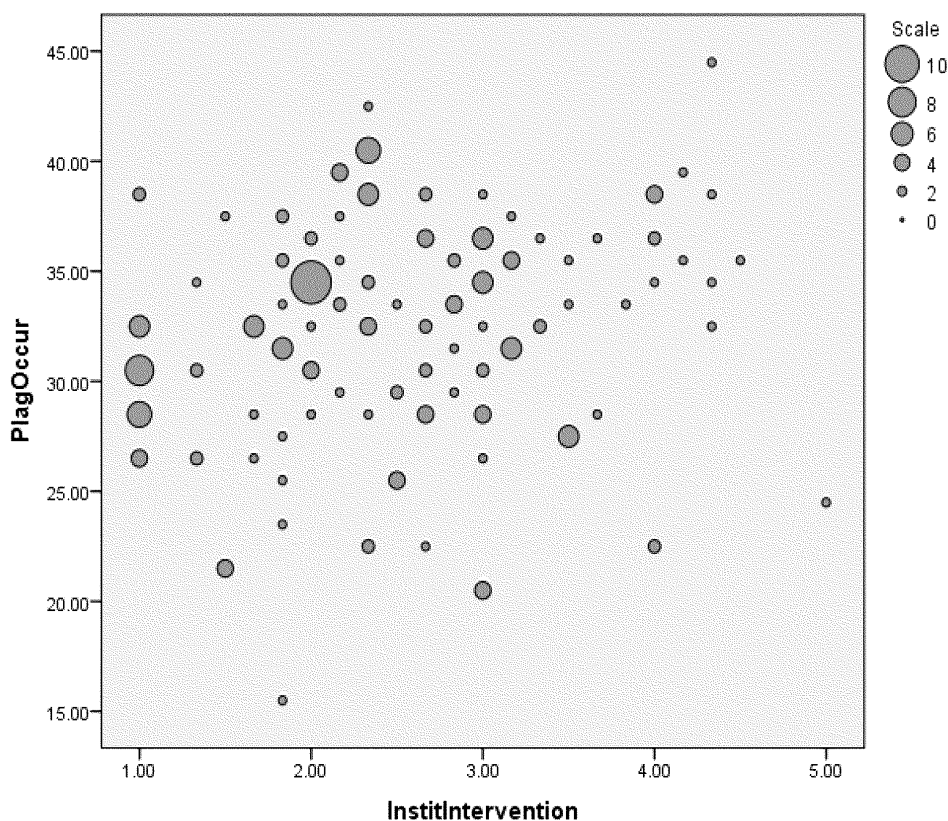


Figure 7. Institution Intervention vs. Student Plagiarism

to verify by the lecturers where the institutions have average student to teacher ratios of approximately 100 : 1 in a number of cases (Udotong, 2012).

The students stated that the most challenging of their problems in adapting in their new study environment was the use of Turnitin. In line with this reasoning, one of the students said that the “knowledge of Turnitin is the beginning of wisdom”. Quite a number of the students expressed “fear” as they understood what “plagiarism” meant, and the system in place to detect it and the penalties that are attached to it in the UK universities.

Some complained bitterly about the “way of writing” and not being able to “...do it as they have been doing it...” They felt they had to learn to write academically and develop the skills (summarizing, paraphrasing, and in-text citation, referencing e.t.c) that UK universities expect them to have mastered at this stage of their learning.

With the “teacher-centered” approach to lecturing in most of the Nigerian universities, the possible lack of student engagement, the possibility that deep-learning has failed to take place, and with the advent of Internet, a cluster of factors allow students

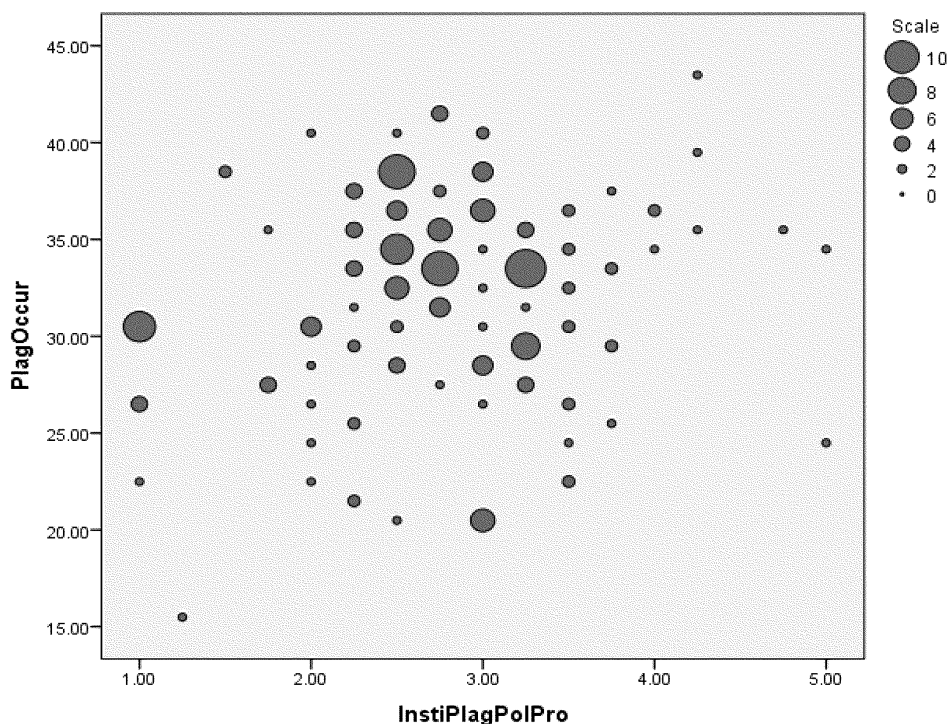


Figure 8. Institutional Plagiarism Policy and Procedure vs. Student Plagiarism

to cut and paste, download and reproduce another person's work without concern, as there are no effective systems in place to detect and penalize them.

For this reason, examination misconducts, falsification of records, academic dishonesty remains one of the major challenges of the Nigerian education system (Olasohinde, 2000; Olasohinde-Williams, Abdullahi & Owolabi, 2003). Their understanding of the UK university study requirement was a combination of skills acquisition, knowledge acquisition, all of which must be demonstrated in their assessments. This requirement in their opinion was enormous and a daunting.

In addition, the presence of an institutional system for the mitigation of student plagiarism, students had to learn to adapt to diverse teaching, learning and assessment styles within the timeframe of coursework deadlines.

Another identified aspect of their challenge they stated was the UK teachers' unfamiliar accent; the method of lecture delivery and the use of virtual learning environments such as Moodle. They found these different from the use of blackboards and the teacher-talking-at-student lecture delivery method which was prevalent in use in their previous Nigerian universities.

Furthermore, they had to learn to research beyond the reading list, which they found to be a challenge because in the past, the material that the teacher presented in class

would have sufficed to make a good grade on the course in their previous Nigerian university.

Some also stated that they also found the effective use of the UK university library a challenge as the Nigerian universities they came from did not have well-equipped libraries; hence they were not used to rigorous search for relevant articles in the library or online.

They had to learn to identify the right materials to use for their research and how to get them. As they progressed in their study in UK, they encountered several learning concepts that they found to be quite different from those they have been exposed to during their past study in Nigeria.

In a bid to learn and perfect the skills of proper academic writing, while trying to get better with their use of the IT Systems, they struggled to cope with these challenges. Some claim that their performance is not usually what they would have had if their previous academic background had been more similar to that which they encountered in the UK.

This impacted on the learning experience of a number of them. Coming from a system where the "result" matters so much, a few of them, speaking retrospectively said that, though they were happy about the exposure and what they learnt, as it was quite different from what they had expected, they may not have paid the amount of money they had paid to come to study as they were not too encouraged by the final results they had.

Recommendations

These recommendations are made with respect to the Nigerian universities and the UK universities. It is recommended in the Nigerian universities that:

- The Nigerian Government through the Federal Ministry of Education should come out with a policy statement on plagiarism and possible penalties. They should have in place guidelines and procedures for their effective mitigation.
- The institutional management staff should ensure that these guidelines and procedures will be used as a consistent way of intervening in the occurrence of student plagiarism.
- Key people should be identified and trained in each department to handle the issues of plagiarism. These should then ensure that others understand what the concept is and be able to manage issues around the occurrence of plagiarism by students.
- Teachers as role model should show good examples worthy of imitation by students and discourage plagiarism in assignments, tests and theses.
- Students should be given handbooks, orientation and re-orientation on the importance of their educational pursuits and on best practices in the use of the Internet, to forestall possible infringement on people's intellectual property.
- The institutions should take up the challenge of preparing the students adequately so they are ready to study anywhere with the appropriate skills.

- Where the Nigerian universities have a course in the first or second year that deals with technical writing, the skills acquired should be re-enforced over their study period.

It is recommended in the UK universities that:

- There should be no assumptions that students coming over to study in the UK universities have the required skills to manage their postgraduate studies effectively without any assistance.
- Since international students may come from a “rote” system, it is good practice to have seminars/tutorials in addition to lectures to clarify the expectations of the new academic context.
- There should be centers for academic writing where support can be provided to the students while they study.

Conclusion

For many reasons, including the lack of an institutional context in Nigeria in which plagiarism is proscribed and effectively punished, as well as other institutional factors such as large, impersonal classes and limited resources, some students in Nigeria succeed by cheating. Students are more engaged with their mobile phones, computers (where the Internet is available), social applications (twitter, Face book, Whatsapp, BBM) than to their academic work. This could be a possible explanation why the student plagiarism culture impacts on their experience when they study elsewhere.

Nigerian universities must create awareness, teach the students the “rules of the game” (Leask, 2006:191) and sustain it over their period of study through an “iterative process” (Ellery, 2008: 507).

The universities will need to *institute* and communicate their policies not only on students and scholars who commit the offense of plagiarism but on the educational environment as a whole.

References

- Arenyeka, L. (2012) Copy, paste and plagiarize. *March 29, 2012*
In Education Available from <http://www.vanguardngr.com/2012/03/copy-paste-and-plagiarize/>
 [accessed 24 January 2013]
- Chiedozie, I., (2012) Don sues CBN governor for plagiarism. *Punch*. April 24. Available from
www.punchng.com/news/don-sues-cbn-governor-for-plagiarism [accessed 14 January 2013]
- Cohen, DV. (1993) Creating and Maintaining Ethical Work Climates, *Business Ethics Quarterly*: Volume 3, Issue 4, October Pages 343–358 DOI: 10.2307/3857283
- David, M. (2012) Effective learning and teaching in UK higher education. Available from
<http://www.tlrp.org/pub/documents/UKHEfinal.pdf> [accessed 12 February 2012]
- Deckert, GD. (1993) A pedagogical response to learned plagiarism among tertiary-level ESL students. *Journal of Second Language Writing*, 2, 94–104.
- Ellery, K. (2008) Undergraduate plagiarism; a pedagogical perspective. *Assessment and Evaluation in Higher Education*, 33(5), pp. 507–516.

- Inegbenebor, B. (2012) Experience of Nigerian students studying abroad. August 9, 2012. In *Education*. Available from <http://www.vanguardngr.com/2012/08/experience-of-nigerian-students-studying-abroad/> [accessed 24 January 2013]
- Kalu, N. (2013) UNICAL dismisses five lecturers. Available from <http://thenationonline.net/new/news-update/unical-dismisses-five-lecturers/> [accessed 15 March 2013]
- Kohlberg, L. (1969) Stage and Sequence: The Cognitive Development Approach to Socialization. In Goslin, DA. (eds) *Handbook of Socialization Theory*. Rand McNally, Chicago #47–480.
- Leask, B. (2006) Plagiarism, cultural diversity and metaphor—implications for academic staff development. *Assessment and Evaluation in Higher Education*, 31(2), pp. 183–199.
- Naija (2012) Plagiarism: 2 Nigerian Teachers Sued, 12 November. Available from www.news.naija.com/14301.html [accessed 15 January 2013]
- NEWS of the PEOPLE (2013) FAREED ZAKARIA who was Disgraced over Plagiarism Storms Nigeria for “Airtel Night of Influence”. Available from <http://www.newsofthepeople-ng.com/?p=821> [accessed 24 January 2013]
- NIGERIANS ABROAD (2011) Indian researcher accused Covenant University professor of plagiarism. March 21, 2011. Available from <http://nigeriansabroadlive.com/indian-researcher-accused-covenant-university-professor-of-plagiarism/> [accessed 19 January 2013]
- Nnabugwu, F. (2012) ICPC beams searchlight on corrupt practices in Nigerian universities. July 25, 2012. Available from <http://www.vanguardngr.com/2012/07/icpc-beams-searchlight-on-corrupt-practices-in-nigerian-universities/> [accessed 24 January 2013]
- Olasehinde, O. (2000) Cheating in Examinations: A contemporary issue in, and a Challenge for Nigerian University education in the new millennium. Change and Choices in the new Century Referred 2000 Conference Proceedings. Pp. 152–159.
- Olasehinde, FAO., Abdullahi, OE., Owolabi, HO., (2003) The Relationships between background variables and cheating tendencies Among students of Instituting Academic Integrity In Nigerian Universities: Psychological Perspectives Of Morality And Motivation 165 a federal university in Nigeria. *Nigerian Journal of Educational Foundations*. 6, (1)
- Oni, B. (1996) The Labour Market and Employment Planning in Nigeria, in Ariyo, A., (ed.), *Economic Reform and Macroeconomic Management in Nigeria*. Ibadan: University of Ibadan Press, pp 339–354
- Oxford Dictionaries. Available from <http://oxforddictionaries.com/definition/english/plagiarism> [accessed 12 February 2013]
- Park, C. (2003) In Other (People's) Words: plagiarism by university students—literature and lessons, *Assessment & Evaluation in Higher Education*, 28(5), pp. 471–488
- Pasternak, S. (nd). Available from http://www.ti-israel.org/_Uploads/dbsAttachedFiles/sigalitpasternak.pdf [accessed 12 February 2013]
- Reddy, MJ. (1979) The conduit metaphor—A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (pp. 284–324). Cambridge: Cambridge University Press
- Trevino, LK. (1986) Ethical Decision-Making in Organizations: A Person-Situation Interactionist Model. *Academy of Management Review*, 11(3), 601–617
- Trevino, LK., Ball, GA. (1992) The Social Implications of Punishing Unethical Behavior: Observers' Cognitive and Affective Reactions. *Journal of Management*, 18: 751–768

- Trevino, LK., Butterfield, KD. & McCabe, DL. (1998) The Ethical Context in Organizations: Influences on Employee Attitudes and Behavior. *Business Ethics Quarterly*, 8(3): 447–476.
- Udotong, W. (2012) in Obinna, C., (2012) Nigeria: Plagiarism, Bane of Nigeria's Educational Devt—Provost. *Vanguard*, 20 September 2012. Available from <http://allafrica.com/stories/201209200813.html> [accessed 18 December 2012]
- Victor, B., Cullen, JB. (1988) The Organizational Bases of Ethical Work Climates. *Administrative Science Quarterly*, 33(1), 101–125.
- Yesufu, TM. (1996) *The Nigerian Economy: Growth without Development*. Benin Social Science Series for Africa

Author Details

Stella-Maris I Orim MSc, MCM, BEng, PMBCS, FHEA

Coventry University, United Kingdom, S.Orim@coventry.ac.uk

Stella-Maris Orim is a current doctoral student and a National Teaching Fellow of the HE Academy. Her research centres on the investigation of student plagiarism in the Nigerian universities from the student, teacher and management staff perspective. She is exploring the possibility of developing a student plagiarism mitigation framework that can be used across the Nigerian universities as a consistent approach to the mitigation of student plagiarism. She has worked as a Research Assistant on the Impact of Policies for Plagiarism in Higher Education across Europe (IPPHEAE) Project and is presently working in Coventry University, UK.

Dr. Erik Borg

Coventry University, United Kingdom, aa3096@coventry.ac.uk

Erik Borg is a senior lecturer at Coventry University's Centre for Academic Writing, where he teaches undergraduate and Master's level modules on writing and research. His research, which centres on research into the nature of writing in a rapidly changing communication environment, investigates particularly intertextuality and multimodal communication. He has published widely, including recent articles in *Assessment and Evaluation in Higher Education*, *The Journal of Writing in Creative Practice*, and *Teaching in Higher Education*.

Dr. (Mrs.) Awala-Ale, I. Isabella PhD, M.L.S, B.Ed, N.C.E

College Librarian, College of Education, Warri, Delta State, Nigeria, idoefemu@yahoo.com
comandidomu_awala@yahoo.com

Dr. (Mrs.) Awala-Ale has worked in both the public and academic sectors of Nigeria. She started her librarianship career with the National Library of Nigeria, Lagos in 1987 and rose to the rank of Librarian 1 in 1990. She then transferred her services to College of Education, Warri as a Senior Librarian in 1993. She became the substantive College Librarian in 2004. She has attended many Conferences and Seminars: Local and International, making several Conference and seminar presentations, published many articles and has to her credit three books on Librarianship.

Copyright © 2013 The authors listed on page 54 grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the author.

INTERNATIONAL MBA STUDENTS' ACADEMIC MALPRACTICE: A QUANTITATIVE SURVEY

Neil Wellman, Julia Fallon

Abstract:

This paper presents the findings of quantitative research conducted between March 2011 and February 2012 with 182, predominantly South-Asian, students drawn from four cohorts of a UK Masters of Business Administration (MBA) programme. It builds upon previous investigations into malpractice amongst international students following the discovery of high incidences of plagiarism and collusion, amongst this student group (Wellman and Fallon 2012). A survey utilised a self-completion questionnaire administered under controlled conditions and aimed to identify students' views on giving and receiving help from others, collaborative working, use of source material and their own, and others' engagement in various malpractices. The findings suggest that individual peer support is largely held to be acceptable, but group collaboration less so but also that a value system is in place which, for instance, endorses support for weaker colleagues but condemns group "free-loaders" and receiving outside support. Respondents generally knew what constituted unacceptable practices and denied engaging in them. However, the reportage of malpractice in others suggests a higher occurrence than self-reportage suggests, with unreferenced copying of source material and group working on individual assignments being most commonly cited. Whilst the data presented is quantitative and we further explore it qualitatively, they will be of value to those working with similar international students, as they highlight some of the attitudinal issues which must be addressed to help such students adapt to and succeed in what for some is an alien social and educational culture.

Introduction

This research stems from our experience at Cardiff School of Management (CSM) in 2008 when 93 MBA students, predominantly from India, were investigated for breaches of academic practices, predominantly plagiarism and collusion. The investigation quickly identified that many saw "copy-pasting" from academic and other sources without references to be acceptable and that seeking and giving help to fellow students was both natural and expected. As it became clear that many students were unprepared for independent study at M-level in the UK, we addressed this issue by re-vamping the initial Induction and Study Skills modules (with considerable effect).

In addition, we decided to research the problem in order to better understand the issues (Fallon & Wellman 2012). The literature, interviews and focus groups suggested that much of what we had called unfair practice (UfP) was often nearer to what Errey (2002), Carroll (2007), and Magyar (2009) observed to be innocents straying into malpractice due to misunderstanding the UK's academic practices and conventions. In particular, we found that students had often only experienced what Timm (2008a) observed to be rote learning with little exposure to independent study and had been assessed primarily by short knowledge based tests and exams. We also found that students often shared their knowledge and expertise to support each other, this ranging from altruistic support and cooperative collaboration to collusion. Apocryphally,

sharing ideas and material was found to be common with one student stating that as they didn't know anything about the subject they went to others who did (with no mention of learning) and another that it was "...right that the seniors" [students from previous cohorts] "*should support the juniors*" (Personal communication 2010).

Research objectives

As a result of the earlier research we decided to further investigate the situation in order to learn more about problems and their causes and to identify remedial strategies. This is an ongoing project which involves both qualitative and quantitative elements.

This paper reports on a quantitative strand with three objectives:

To identify students':

1. Study and assignment preparation methods
2. Views regarding various practices regarding giving and receiving help, working in groups and use of source material
3. Views and practices regarding students own and others' engagement in poor academic practices

This paper reports on survey findings in relation to objectives 2 and 3, drawing from findings from four samples of MBA students, totalling 182, during 2011 and 2012. The survey is part of an ongoing project and it is intended to administer it on a regular basis to all MBA cohorts as part of the induction process.

Literature Review

There is an extensive body of literature covering the topics under consideration and, some of the most informative are included here. These cover research into cultural and educational differences and their experiences of studying in the UK. Leading contributors include, Hayes and Introna (2005), Bennett (2005), Handa and Power (2005) and Timm (2008a) all of whom have written about South Asian students in particular. Other useful insights have been found within Bailey (2006) Campbell-Evans and Legget (2007), McCabe, Banwell (2003), Dukerich and Dutton (1993) and Flynn (2003) who all report similar issues with other cultures.

Interestingly, Banwell (2003) reports on how some cultural values encourage peer support, a phenomenon also noted by Timm (2008a) and Hayes, Introna and Whitley (2006). Whilst seeking others' help on an assignment may be valuable and encouraged, Barrett & Cox (2005) state, the line between collaboration and collusion is unclear.

The problems faced by students adapting to the UK (and other Western) styles of education are discussed by many authors and a major difficulty is often with writing and language. This has been explored by a number of authors including most notably Errey (2000), Carroll and Appleton (2001), Barret & Malcolm (2006), Hayes and Introna (2006) and Carroll (2007). One coping strategy, coined "*patch-writing*" by Howard, is the practice of loosely linking together sections of text drawn from disparate sources into what Dryden describes as a "*beautiful patchwork*" (cited in Introna and Hayes (2005:4), or has also been called by Barrett and Malcolm (2006) an "*incoherent*

whole.” However, Introna and Hayes (2005) endorse Shi’s earlier view that such “*grey plagiarism*,” is a natural part of international students’ development whilst Schmitt (2007), Carroll (2007), Magyar (2009) and others, express concern that this may lead to accusations of plagiarism.

The predicament is, as Yeo (2006) notes, in the interpretation of line between paraphrase and plagiarism and as Alam (2004) suggests, this is particularly so for international students. Many authors suggest this is due to students’ past educational experience with, for India, Banwell (2003) Hayes, Whitley and Introna (2006) and Timm (2008a) reporting that there are often rote taught courses with in the main, reliance on a single text and little requirement to read more widely. Plagiarism is often not a concern, observe both Park (2003) and Timm (2008a) whilst among Chinese students Introna and Hayes (2005), expressed the view that repeating respected academics’ words is expected.

Nevertheless, Park (2003), Johnston (2003) and Clark (2008) fear that intentional malpractice exists in the form of inadvertent plagiarism but that this cannot be the case for collusion (Johnston 2003). Nor perhaps is purchasing assignments accidental, for as Lancaster and Clarke (2009) state that assessments are often purchased from commercial websites/essay mills or obtained from current or past students, family and friends. Indeed, Introna & Hayes (2005) discovered that 60% of Asian students admitted to preparing work for others and 60% admitted to having submitted other’s work..

There are many previous surveys of both domestic and international students’ views of and participation in various potential malpractices. Useful examples include the work of Brown and co-authors (See Brown 1999, Brown & Choong 2005, Brown & Weible 2006 and Brown & McNery 2008). These demonstrate a large body of research about the occurrence, attitudes and motivations regarding academic dishonesty amongst a number of populations, spread over two decades. Consequently, Part D of this survey’s questionnaire is based on Brown’s.

Other influential sources include Miller, Shoptaugh and Parkerson (2008) and Megehee and Spake (2008) whose research identified underreporting of malpractice by respondents. Butterfield and Skaggs (1988) also suggest that asking for the believed transgression rate of other students help obviate this problem; a solution incorporated into the survey.

Others consulted and provided useful insights include: Gururajan and Roberts (2005) Barrett and Cox (2005) McCabe, Dukerich and Dutton (1993) and Moon (1999) whose work investigated attitudinal issues using scenarios posing moral dilemmas. Also useful were Yeo (2006) and Iyer and Eastman (2006), who discussed the effects of social and peer influence. Other significant influencers for their discussions about academic dishonesty were Wilson (2008) Swift and Norris (2008), Dawson and Overfield (2006) Darbyshire and Burgess (2006), Tennant and Duggen (2008) and Etter, Cramer & Finn (2006).

Research Method

A quantitative approach was adopted using, as others have done (see literature section) and as Miller, Shoptaugh and Parkerson (2000) advise, a survey to gather statistical benchmark data. As recommended by Saunders, Lewis and Thornhill (2009), Bryman and Bell (2007) and Jankowicz (2007), parallel strands utilising quantitative techniques will explore the findings to gain greater insights and understanding.

The survey instrument

The survey was a four part questionnaire with Part A gathering demographic and assessment experience data and Part B data on study and assignment practices. However, this paper focuses on the findings for:

- **Part C:** This aims to gather data on giving and receiving help, working in groups and use of source material. A five point Likert scale measures agreement with 14 statements.

Whilst the initial questionnaire (version Ca) also included tick-box questions asking whether respondents and others had participated in each practice, they were later removed. The reason for this was because it was felt that the survey should focus on gathering general opinions and due to feedback regarding the questionnaire's complexity and length (version Cb).

- **Part D:** This aims to identify the frequency with which students and others had engaged in, and their views on the efficacy of, various practices, again using Likert scales (the first of six, the second five). Questions were adapted from Brown's (1999) survey but with the addition of one asking whether respondents believed others engaged in each practice.

The survey is presented (and titled) as investigating MBA study practices with an introduction explaining that it is anonymous, that individuals may opt-out and that it aims to inform the School's study skills programme. Similar to Megehee and Spake (2008), questions (other than the last two) avoid using pejorative terms like the words dishonest which were seen as problematic because they would influence answers. Instead, all were simple statements of behaviour, albeit some blatantly unacceptable, with no attached value judgements.

Megehee and Spake (2008) suggest that respondents may be unwilling to incriminate themselves by admitting to malpractices and cite Brass et al's (1988) opinion that reportage of *others'* behaviour is a better indicator of the *individual's* than self-reportage. In support, they cite Zey-Ferrell et al, (1979); Erikson,(1988); Underwood and Szabo, (2003) and other's work on social contagion theory which suggests that, as individuals tend to both conform to and reflect group norms, this is a good proxy for their actual behaviour. This technique was also utilised by others, including Chapman and Weiss (2000) and Miller et al (2008) who suggest that otherwise, under-reportage occurs.

The questionnaire was piloted in January 2011 with 21 students following which some minor adjustments to wording were made (findings are not included in this paper). To ensure that questions were clear and to give guidance, a PowerPoint presentation was used to project each question with responses recorded on a simplified answer sheet.

This is the preferred method of administration although hard-copy paper versions have also been used when necessary.

Administering the survey and data analysis

Convenience sampling was taken, based on the opportunity to administer the survey to students within plenary sessions. To date, the survey has been administered with four samples, as:

- March 2011, 84 students: administered en-route from Cardiff to Heathrow airport at the start of a field trip to Budapest (paper version Ca)
- April 2011, 22 students: administered in Budapest during a second trip (paper version Ca)
- July 2011: 55 students: administered during a plenary session (PowerPoint version Cb)
- February 2012: 22 students: administered during a plenary session (PowerPoint version Cb)

Coded answers were entered into individual sheets within a Microsoft Excel database. Whilst each sample and demographic measure is identifiable, thus facilitating correlation analysis, this paper presents only summary data. Excel utilities were used to compute totals, means, indices etc. and to generate tables of findings and charts. Initial findings, based on the first two samples, were presented at two HEA seminars in 2010 where they aroused considerable interest and positive comment.

Research limitations

It is acknowledged that the method may be criticised in several ways. Firstly, the sample, at 182, is relatively small. Nevertheless, this represents 9.6% of the 1894 full-time MBA students for the period surveyed; thus suggesting a confidence interval of 6.5% at the 95% level (<http://www.surveysystem.com/sscalc.htm> retrieved on 11 February 2013). Similarly, the sample is skewed toward young adult male South Asians who have not previously studied in the UK. However, this accurately reflects the, then, MBA student population (the emphasis on South Asians has since reduced).

Secondly, if as Penshaw, Straughton and Albers-Millers (2001) (cited in Megehee and Spake 2008) suggest, students tend to overestimate others' malpractice, this would infer that such data may be exaggerated. Nevertheless, we feel that such differences, whilst being treated with caution, are significant.

Thirdly, some questions appear ambiguous (e.g. that others may help "...in any way they can"). However, this is intentional as we aim to measure broad perceptions and the propensity for established behaviours to lead to transgression rather than to admission of fault, and, as Lupton, Chapman and Weiss (2000) suggest, rely on students' interpretation of the acceptability of practices.

Finally, that questionnaire results may not be reliable, especially as there may be a tendency to tick Likert scales "down the middle" or randomly. However, the

Demographic profile of the sample

Nationality/Country		%
South Asian	163	89.6%
China	6	3.3%
Other Asia	0	0.0%
Middle East	2	1.1%
Africa	6	3.3%
USA/Canada	4	2.2%
S America	0	0.0%
East Euro	0	0.0%
EU/ West Euro	1	0.5%
UK	0	0.0%
Othr	0	0.0%

182

Age		%
< 20	0	0.0%
20-23	102	56.4%
24-27	67	37.0%
28-31	8	4.4%
32-35	2	1.1%
36-39	0	0.0%
> 40	2	1.1%

181

Male		%
Male	138	75.8%
Female	44	24.2%

182

Award		%
BA	47	27.6%
BSc	26	15.3%
Othr UG	37	21.8%
MA	3	1.8%
MSc	4	2.4%
Othr M	2	1.2%
PhD	0	0.0%
Othr	51	30.0%

170

Country of Study		%
India/Pak	155	86.6%
China	3	1.7%
Other Asia	0	0.0%
Mid East	2	1.1%
Africa	6	3.4%
USA/Can	4	2.2%
S America	0	0.0%
East Euro	0	0.0%
EU/West Eu	1	0.6%
UK	8	4.5%
Othr	0	0.0%

179

Subject		%
Bus Mgt	61	35.7%
Commerce	11	6.4%
Fin A/Cs	13	7.6%
Othr Bus	1	0.6%
Othr Soc Sc	19	11.1%
IT- Comp St	19	11.1%
Science	15	8.8%
Eng Sc	32	18.7%

171

overall n= 182

The majority (64.7%), held BAs or BScs with some holding others, such as BCom, a common Indian award. Others (30%) held unspecified higher awards and a few (5.4%) M-level qualifications; none held doctorates. Half (50.3%) studied business, management or related subjects and a further 11.1% other social sciences. Similarly, 11.1% had studied IT and the remainder (27.5%) science subjects with a significant number (18.7%) having studied engineering (again, popular in South-Asia). Such a profile is again typical of CSM's MBA population with many, as Quality Assurance Agency for Higher Education 2007 Benchmarks (<http://www.qaa.ac.uk/Publications/InformationAndGuidance/Pages/Subject-benchmark-statement>

-Masters-degrees-in-business-and-management.aspx accessed 16 May 2013) recognise, seeing it as building upon their previous award, whether business related or not. However, the substantial minority who have not previously studied business, or other social sciences, are likely to experience difficulties in moving between disciplines and one can empathise with a South-Asian engineering graduate coming to terms, not only with a new country, culture (an weather!), but also with new pedagogical approaches and academic expectations.

Part C: Views on and participation in potential malpractice

This section asks 14 questions regarding views and engagement in a range of practices which, at their extreme, may be construed as malpractices. Statements were asked about three sets of practices, (although “scrambled” within the questionnaire), as:

- receiving and giving help for individual assignments;
- working in groups on individual assignments;
- using source material.

Three questions were asked in relation to each statement:

- A tick-box to indicate whether the respondent had done this;
- Whether the practice was acceptable (“OK”), utilising a five point Likert style scale (1= very much agree to 5 = very much disagree).
- A tick-box to indicate whether the respondent believed that others do this.

Note: Questionnaire version Ca, including all three questions, was administered to the first two samples ($n = 106$); subsequent samples used version Cb ($n = 77$), with only the scale question.

As noted, statements were ambiguous in asking whether others help “...in any way they can”. Whilst this may lack precision, our aim is to identify tendency or propensity rather than answers to blunt questions such as “do you cheat?”, thus inevitably inviting the answer “no”. Furthermore the response scale is sufficient wide to suggest that replying “very much agree” suggests a very different attitude from “very much disagree”. We therefore maintain that a low mean or index allows for nuances of interpretation and is thus sufficient to suggest broad agreement and vice versa.

Appendix A shows summary data with the key findings discussed below. Whilst the tables contain raw data the charts below have been adapted to allow more direct comparison of the data; notably the “OK” scale is reversed (from 1–5 to 5–1) to express the level of agreement with the statement and the means indexed to $x/100$. As the “very much disagree” ranking thus equal 1, a mean of 1 or an index of 20 would therefore denote complete disagreement, a mean of 3 or index of 60 the mid-point and 5 or 100 total agreement. For the purposes of this discussion 40 (mean 2) is taken to be the critical threshold. The reportage of self and others undertaking the practices are percentages.

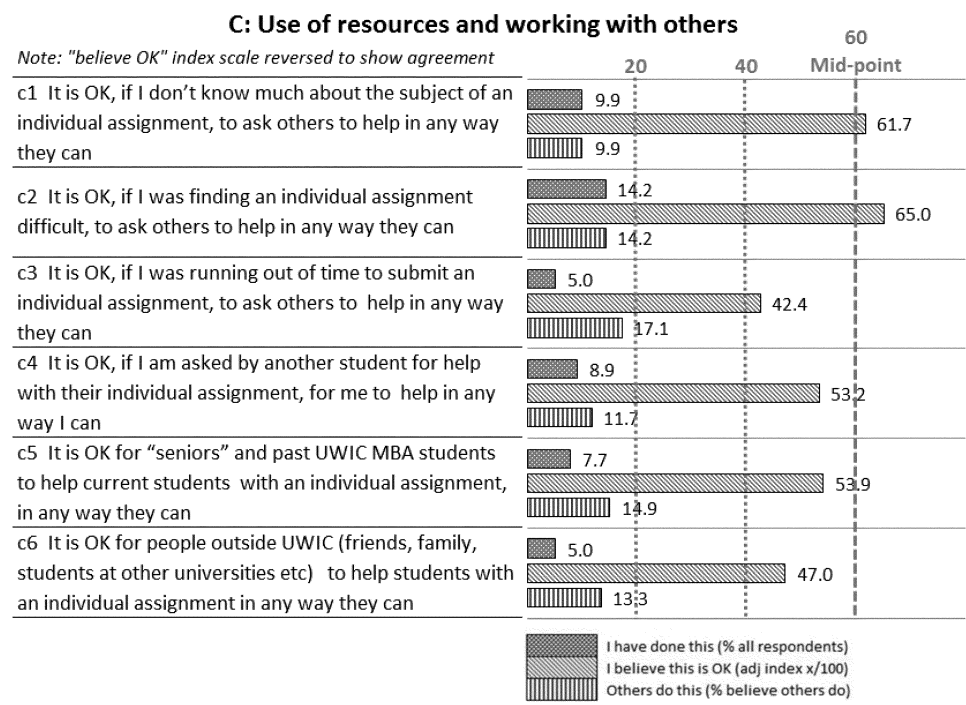


Figure 1. Receiving and giving help on an individual assignment (index x/100 and %

Receiving and giving help on individual assignments

Six value-free statements (c1–6) were given about working with others on an individual assignment (see, Figure 1 for indexed scores).

As the level of acceptability for all types of peer support score above the 40 threshold we infer that it is many who consider it acceptable to get from, or give help to, others. This appears particularly so when little is known about the subject (c1: 61.7) or it is found difficult (c2: 65.0), both of which exceeded the 60 mid-point (the highest indices in Part C). Acceptability is slightly lower for receiving help from seniors, giving help and receiving help from outsiders (c5: 53.9; c4: 53.2; c6: 47.0) and to just above the 40 threshold if due to running out of time (c3: 42.4). With a range of 42.4–61.7 and a mean of 53.9), the findings tend to support our own and the literature’s observation that South-Asian students expect to give and receive peer support.

Self-reportage rates display a varied pattern (range 5.0–14.2%, mean=8.45%) with receiving help due to difficulty, poor knowledge or from seniors, and giving help when asked again being the most frequently reported (14.2%, 9.9%, 7.7%, 8.9%). Receiving help due to time-management from outsiders again scored less (both 5%). As one may expect from Butterfield and Skaggs (1988), the reportage of practices by others, is generally higher (range of 5.0–17.1%; mean = 13.5%) although the increase for some is

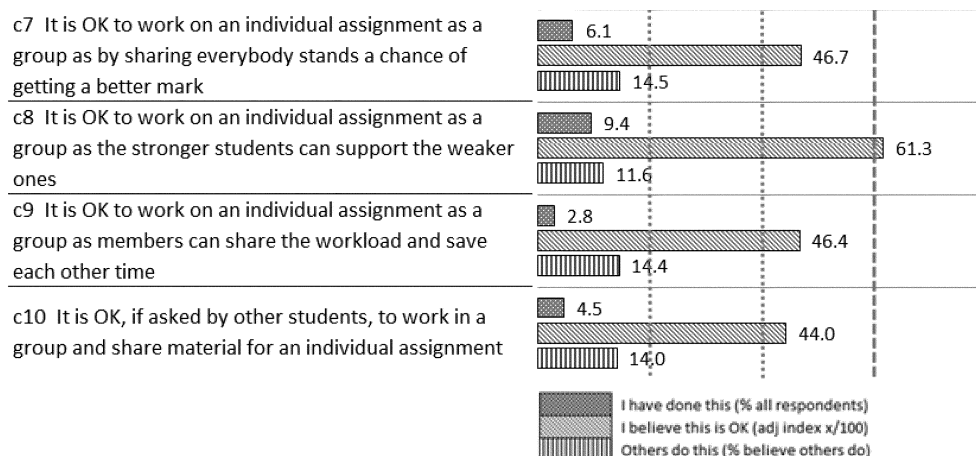


Figure 2. Working in a group on an individual assignment (index $x/100$ and %

quite low. Indeed reportage due to for difficulty and poor knowledge remained the same (checks confirm that this is not a recording error). In contrast, reportage of seeking help due to time constraints increased by a factor of $x3.42$ and that for gaining help from outsiders by $x2.66$, perhaps reinforcing Brass et al's (1988) thesis.

The low acceptability indices and self-reportage rates for these two statements (c3, c6) contrasts with their relatively high rate of reportage in others. This is reinforced by inspection of the data (Appendix A) where they jointly have the greatest skews towards unacceptability. This perhaps indicates that, whilst felt to be least acceptable, they nevertheless occur, and if Butterfield and Skaggs are correct, at rates nearer 13.3% and 17.1% than 5%. Such findings are intriguing and may lead one to suspect that some value judgement is being made about these practices being less acceptable than the others.

Whilst we do not suggest that the data are conclusive, they nevertheless support the proposition that peer collaboration is acceptable to a significant portion of the sample with a small but significant number reporting that they and others have done so. It is notable that the two statements receiving greatest support (c1; c2) may be interpreted as reflecting an ethic of mutuality and altruism.

Working in groups on individual assignments

This section comprised four statements concerning group working on individual assignments.

With a range of 44.0–61.3 (mean = 49.6), group working is apparently less acceptable than individual collaboration, the only exception being when it provides support to weaker colleagues (c8: 61.3). By comparison, all other practices received relatively low support, little above the 40 threshold and broadly equivalent to the indices for c3 and c6.

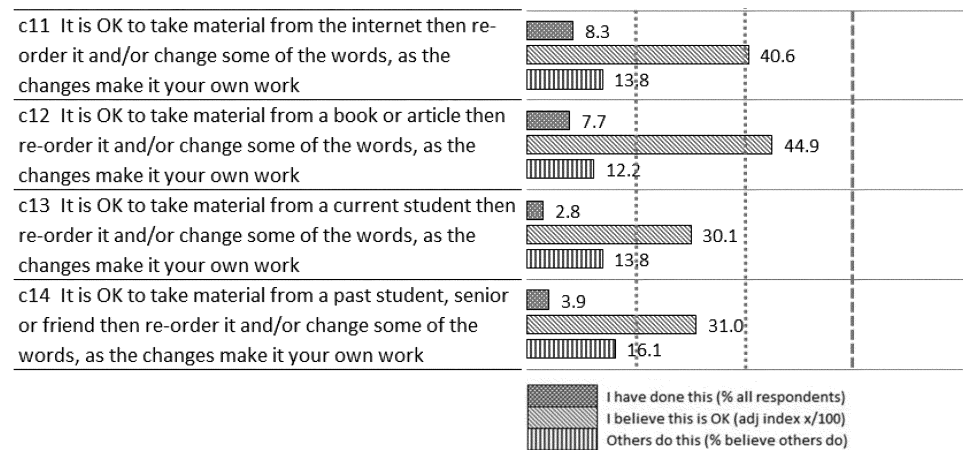


Figure 3. Use of source material with re-ordering and/or changes (index $x/100$ and %

Self-reportage was again low, ranging from 2.8% to 9.4% (mean = 5.7%), with c8 again reflecting the relative acceptability of supporting weaker colleagues. Despite this, rates of reportage of others increased by factors of $x1.23$ to $x5.14$ (range of 11.6–14.5%; mean = 13.6%), thus bringing them into the same range as those for c1–6. Whether the greater differential between the rates for sharing the workload and saving time (c9: 2.8% vs 14.4%) reflects similar differentials for getting help due to time pressure (c3) is unclear but it may be significant that both relate to poor time management or low commitment and are perhaps perceived as undeserving of aid.

Overall, the data suggest only low levels of support for group collaboration, although small but significant numbers report that it does happen. However, the relatively higher acceptability and self-admission for group-working to support the weaker, if taken with c2 and c1 above, again suggests a greater acceptance of such acts of peer-support as opposed to the more instrumental practices of time-saving or gaining better marks (c3; c6; c9).

Use of source material

This comprised four statements regarding whether source material is re-ordered and/or worded to “... make it your own work.”

The results suggest that acceptability is low, ranging between 30.1–44.9 (mean = 36.7) with only re-working internet (c11) and text sources (c12) reaching the 40 threshold and adapting work from current (c13) or past (c14) students being significantly below it. This strength of feeling is borne out by the numbers who felt such practices to be totally unacceptable (see Appendix A), where for the latter 44.8% and 67.8% respectively answered very much “disagree” for them being “OK”.

Self-reportage ranged from 2.8% to 8.3% (mean = 5.7%) with only adapting material from internet and text sources (c11: 8.3%; c12: 7.7%) falling into the mid-range of self-reportage. Nevertheless, all practices were felt to be significantly more common in

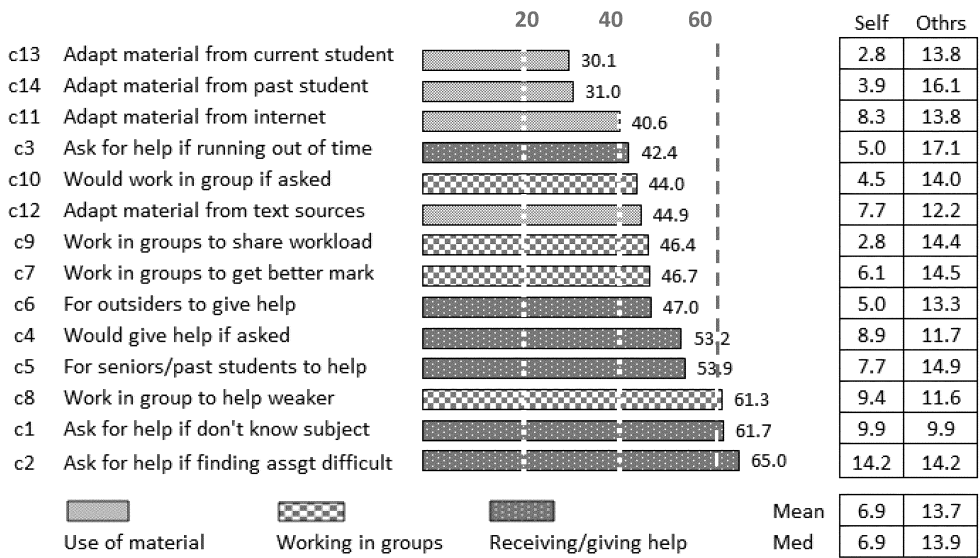


Figure 4. Ranking of practices by perceived acceptability (index $x/100$; $n = 182$)

others, with a range of 12.2% to 16.1% (mean = 14.0%) with the adapting past students' work (c14: 16.1%) receiving the second highest overall score (after c3's 17.1%).

That these four practices were amongst the six felt to be least acceptable (see Figure 4) is perhaps unsurprising since the dangers of incorrect use of sources and poor referencing have been continually stressed. However, significance may be attached to the unusually high self vs others differentials ($x_{4.13}$ and $x_{4.93}$) for adapting the work of others (c13/14) which, if Butterfield and Skaggs are correct, suggests that the rates of this are higher than self-reportage suggests.

Ranking of practices' acceptability

Figure 4 shows all practices ranked by their index together with the percentages for self-reportage, reportage of others and the differential between the two ($S \div O$).

As can be seen, most practices rate above the 40 threshold with only two (c13/14) falling below it and three (c8/1/2) above 60. However, there are clear groupings, with practices relating to plagiarism least acceptable (c13/14/11), working in groups taking centre-field (c10/9/7) and receiving and giving individual help the most acceptable (c4/5/1/2). Major variations to this pattern are, asking for help due to running out of time (c3) which is lower than others in its category and working in a group to help weaker students (c8) which is higher. It is noticeable that the five most acceptable practices (c2/1/8/4) all relate to collaboration, perhaps again supporting the view that peer-support is perceived as a natural and desirable trait.

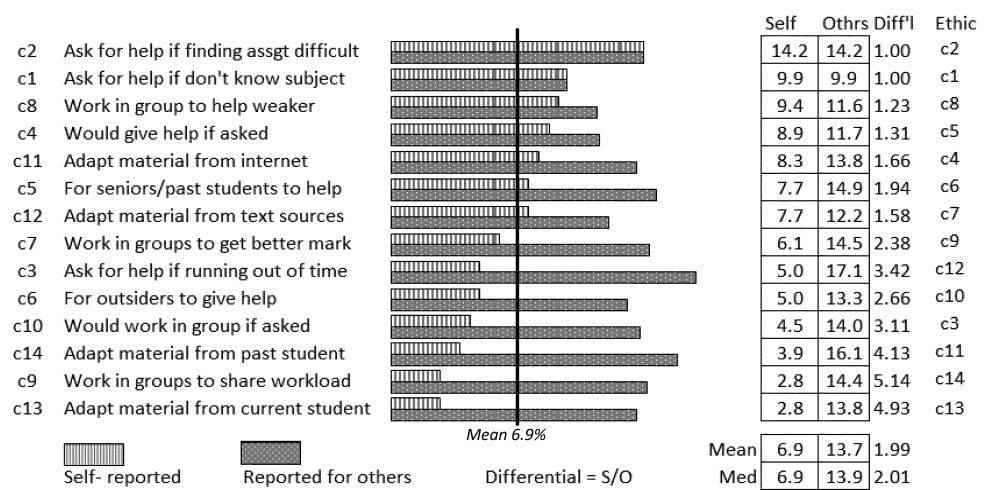


Figure 5. Ranking of practices by self-reportage (% stating self/others do it; n = 106

Ranking of practices by reportage

The Figures below compare self-reportage with reportage by others, Figure 5 ranked by the former and Figure 6 the latter, with acceptability ranking on the right.

Self reportage ranged from 2.8% to 14.2% (mean = 6.9%) along a fairly constant gradient which broadly mirrors practices' acceptability, with five of the top seven (c1/2/4/5/8) are also the most acceptable. Whilst unsurprising respondents more readily admit to what they perceive as more acceptable, there are some anomalies. Notably, adapting material from the internet (c11) is seven points higher than its acceptability ranking and working in groups to share the workload (c9) is six lower.

Ranking by reportage of others shows greater rates, ranging from 9.9% to 17.1% but with less variation around the mean of 13.7%. However, they fail to correlate with acceptability, notable variations being that practices felt to be least ethical (c3/14) are reported as relatively common and others felt to be more acceptable (c1/4/8) less so. Furthermore, in what seems a reversal of Brass et al's (1988) hypotheses, the three least reported practices (c4/8/1) are amongst the most commonly self-reported, and in the upper range of what is deemed acceptable.

Whilst the sample is relatively low (106 compared with 182 for the rankings), the data suggest that all practices occur, and if one is to believe Brass et al, at levels between 10% to 17% rather than 3% to 14%. Furthermore, as Figure 5 shows, self-reportage rates both vary more than that for others (SD 3.071, mean=6.9 vs SD 1.687, mean 13.7) and that they more closely matches overall perceptions of acceptability.

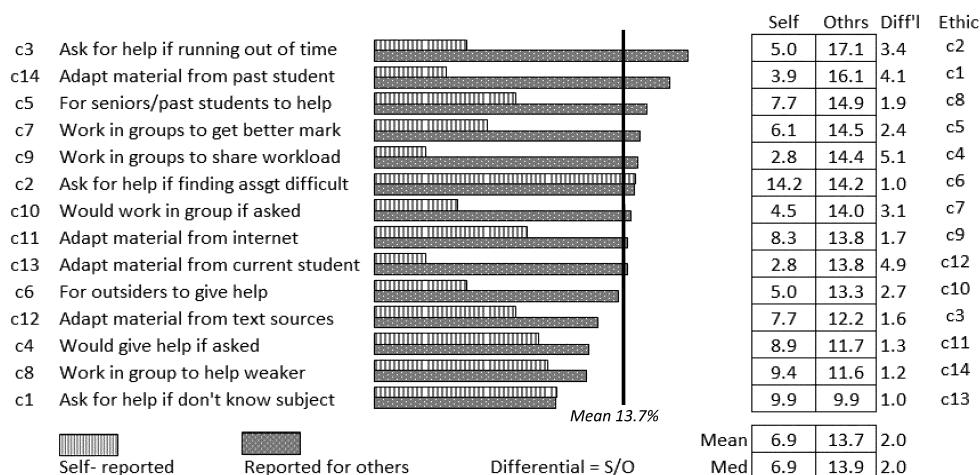


Figure 6. Ranking of practices by reportage for others (% stating self/others do it; n = 106)

Part D: Participation and views on unfair practices

Part D contained questions aiming to elicit students' views and participation in ten unfair practices. These were based on Brown's (1999) 16 questions but were modified to adapt them to the MBA context. In particular, his questions regarding examinations were deleted as they do not apply and questions 5/8/9 regarding the use of social network sites and paper-mills and online translation were added to reflect current practices.

Three questions were asked in relation to each practice:

- Whether the respondent had carried out the practice, utilising Brown's six point Likert scale (1= frequently to 6 = never);
- Whether the practice was acceptable ("OK"), again utilising Brown's five point Likert style scale (1= not at all OK to 5 = perfectly OK);
- A tick-box to indicate whether the respondent believed that others do this.

Appendix B gives summary data and the main findings are discussed below. As with Part C, data used in the charts below have been adapted by the scales being indexed to x/100 and the self-reportage scale reversed and adjusted from 1–6 to 1–5 to make the indices comparable with others. Thus, as in Part C, a "score" of 20 denotes complete disagreement 60 the mid-point and 100 total agreement; again, 40 is taken as the critical threshold. The reportage of others is a percentage of respondents believing that others participate in such practices.

Figure 6 gives summary findings for each question.

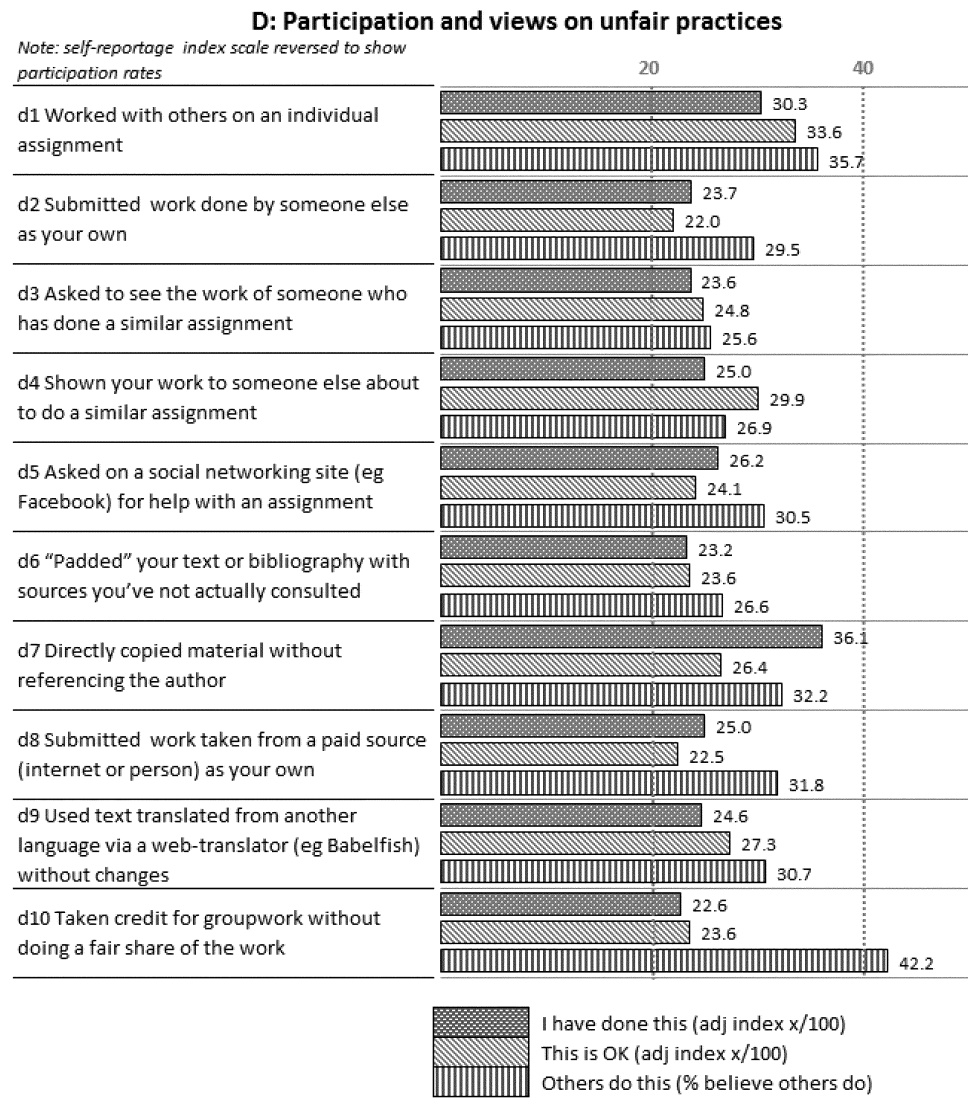


Figure 7. Participation and views on unfair practices (index x/100 and %; n = 182)

Practices’ acceptability

Whilst the practices do not directly match those in Part C, there are sufficient similarities to make comparison worthwhile. Firstly, the “OK” indices are lower (range 22.0–33.6, mean = 25.8: cf. Part C mean = 47.7). Why this is so is unclear although many questions express unequivocally unacceptable, practices (compared with the ambiguity of Part C) and thus more readily attract low scores.

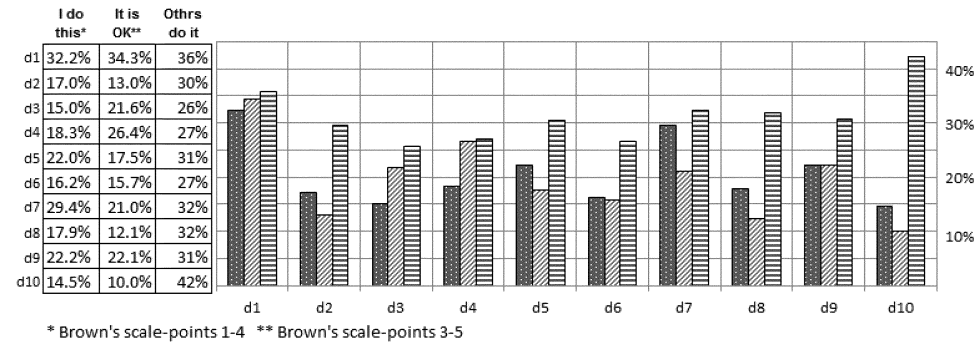


Figure 8. Support for statements expressed as % of sample (%; n = 182

Amongst these, submitting work which is not one’s own (d2: 22; d8: 22.5) and poor use of sources (d6; 23.6; d7: 23.4) received the lowest scores along with taking undue credit for group-work (d10: 23.6) which perhaps represents the dislike of “free-loading” as the flip-side of collaboration.

The highest scores were for working with others (d1:33.6) and showing work to others (d4: 29.9) were the most acceptable, it perhaps significant that, as with the higher scores in Part C, they relate to collaborative working. Otherwise, asking to see others’ work or for help via social networks received low ratings, perhaps reflecting low ratings for similar questions in Part C (c13/14).

Using translation software (d9) merits explanation as it relates to Jones’ (2009) “back-translation”; where e-translation converts source material to a second language then back again to create text which is untraceable by plagiarism software such as Turnitin (and often unintelligible). Whilst this was judged to be the third most acceptable practice (27.3) and we have encountered back-translation on the MBA, we are also aware of students drafting material in their own language and then e-translating it to English. As the question could be construed to mean either, we do not attach significance to these data.

Reportage of practices

In contrast to Part C, the indices of self-reportage of practices (reversed to show participation and adjusted to a scale of 1–5) are relatively high with a range of 22.9–30.3 and a mean of 26 with all close to or exceeding their acceptability rating. Even discounting Brown’s (1999) last two scale-points (5–6) still results in 20.4% acknowledging that they have undertaken such practices more often than “rarely”. Reducing the criteria still further by measuring just 1–3 equates to 14.6%. Thus, even with such liberal interpretations, self-reportage substantially exceeds Part C’s mean of 6.9% and maximum of 14.2%. Figure 7 shows similar computations for other questions (taking self-reportage at 1–4 and acceptability at 3–5).

The higher scores for self-reportage are of interest as whilst working with others (d1) rates highly (30.3: 32.2%) the highest index is for unreferenced copying (d7) at 36.1, although due to its marks distribution it rates only 29.4% in Figure 7. Whilst all other indices are above the 20 mark, none are significantly so and none reach the 40 threshold. However, reference to Figure 7 shows that significant portions of the sample felt that some practices are acceptable (at the 3–5 level), most notably asking to see other's work (d3: 21.6%), showing others one own work (d4: 26.4%) and using text translation software (d9: 22.1%).

As the reportage of other scores represent straight percentages, the thresholds of 20 and 40 are not relevant. It is therefore clear that a substantial number feel that others do act in the ways given with all being above reportage rates in Part C where none scored above 13.7%. The most common are taking undue credit within group-work (d10: 42.2%) and working with others (d1: 35.7%). Other scores over 30% were for seeking for help on a social network (d5: 30.5%), copying without referencing (d7: 32.2%) and using online translation (d9: 30.7%).

As Figure 7 clearly shows, working with others (d1) stands out with all measures above 30% perhaps corroborating the earlier evidence that this is felt to be generally acceptable. That the second highest self-reported practice is copying material without referencing (d7) is surprisingly, not for being common, but rather for the candour of respondents. That freeloading in groups (d10) is felt to be unacceptable but common is again unsurprising; what is surprising is that it receives slightly higher ratings than purchasing work (d8) for both acceptability and self-reportage.

Comparison of findings with Brown's surveys

Although Brown's survey asked some different questions with USA samples, it is valuable to compare his and our findings for similar questions. Figure 8 does this, showing the means from the various surveys reported by Brown et al (1999, 2005, 2006, 2008) compared with questions from our surveys (utilising Brown's methodology of counting all who admitted to each practice [1–5] with the "never" point [6] excluded so that low means indicate higher rates of occurrence).

Differences range from -0.75 to $+0.19$ and are therefore not great, and largely reflect the range of means found within Brown's own surveys. Whether significance can be attributed to the fact that most results are marginally below Brown's (indicating higher occurrence) is moot. Whilst statistical tests may be applied (e.g. chi2), we contend that the sample is as yet insufficient to support this, and even if it did, the differences are not substantial. Further sampling may prove more conclusive.

Conclusions

We have presented a large volume of data, not all of it internally consistent or comparable. Notably, reportage of all factors within Part D is significantly higher than in Part C. Nevertheless, we feel that the data, combined with experience of working with and interviewing students, enables us to draw several broad conclusions.

Note: a low mean denotes more frequent occurrence

	CSM	B av	Diff'I
d1 Worked with others on an individual assignment	3.67	3.48	0.19
d2 Submitted work done by someone else as your own	3.53	4.24	-0.70
d3 Asked to see the work of someone who has done a similar assignment	4.17		
d4 Shown your work to someone else about to do a similar assignment	4.15		
d5 Asked on a social networking site (eg Facebook) for help with an assignment	3.40		
d6 "Padded" your text or bibliography with sources you've not actually consulted	3.96	3.85	0.11
d7 Directly copied material without referencing the author	3.32	4.05	-0.73
d8 Submitted work taken from a paid source (internet or person) as your own	3.49	4.24	-0.75
d9 Used text translated from another language via a web-translator (eg Babelfish) without changes	3.23		
d10 Taken credit for groupwork without doing a fair share of the work	4.12	4.35	-0.23

Figure 9. Comparison of means with Brown's surveys (mean of those having done practice x/3)

- Working with peers is largely considered acceptable but with individual peer-support being more so than working in groups. Whilst Part C (c1–6; c7–10) suggests relatively low occurrences of both practices, Part D (d1/3/4) suggest that it is in fact more common. This finding largely corroborates the literature and our own qualitative research.
- There is some evidence to suggest that gaining help due to time constraints, from outsiders (c3/6) and freeloading within groups (d10) are held to be unacceptable whilst joint working to support weaker students or due to low subject knowledge (c1/2) are more so. This suggests that value judgements are at play with the former being viewed as vices and the latter as virtues.
- Part C (c11–13) suggests that students are aware of the unacceptability of using source material incorrectly and few admit to having done so although they believe it more common in others. However, Part D less suggests a higher degree of acceptability and significantly more occurrences by all parties. This variations cannot easily be accounted for and merit further investigation.
- Whilst Part C found most practices to be more common in others than respondents (means 6.9% vs 13.7%), Part D reported both higher rates of each (means 20.5% vs 34.4% at the 1–4 level) and also much narrower gaps, with d7, copying without referencing higher. Whilst the variations require exploration, if Butterfield and Skaggs are correct, the practices shown in Table 2 are the most and least prevalent.

However, we acknowledge several caveats. First, as the research is quantitative it provides only a surface impression. Secondly, that the statements in Part C contain ambiguities which make interpretation problematic. Both will be addressed within further qualitative research, notably focus groups, to explore the issues and gain greater understanding. Thirdly, the samples are relatively small: and they are skewed to young adult, male South-Asians. These we accept but the ongoing programme of surveying will build the sample and, due to the gradual widening of our MBA recruitment, gradually resolve issues of skew. Finally, we only surveyed MBA students. Whilst we do

not currently intend to extend our research to other populations we invite colleagues who wish to do so to contact us.

References

- Alam, LS. (2004) Is plagiarism more prevalent in some form of assessment than others? In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 49–57). Perth, 5–8 December. Retrieved on 24 April 2013 from http://www.utas.edu.au/_data/assets/pdf_file/0003/167376/Hrasky.pdf
- Bailey, C. (2006) Comparing the experience of Chinese and West African students at a British university: findings from a survey. Retrieved <http://wlv.openrepository.com/wlv/handle/2436/98516>
- Banwell, J. (2003) Chinese and South East Asian students' perceptions of plagiarism and collusion, Northumbria University, Retrieved 24 April 2013 from <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=PLAGIARISM;356ad9c4.04>
- Barrett, R., & Cox, AL. (2005). At least they're learning: The hazy line between collaboration and collusion. *Assessment and Evaluation in Higher Education*, 30(2), 107–122.
- Barrett, R. & Malcolm, J. (2006) Embedding plagiarism education in the assessment process. Retrieved January 1 2010 from <http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/23/18>
- Bennett, R. (2005) Factors associated with student plagiarism in a post-1992 university. *Assessment & Evaluation in Higher Education*, 30(2), 137–162.
- Brass, DJ., Butterfield, KD, & Skaggs, BC. (1998), Relationships and Unethical Behavior: A Social Network Perspective, *Academy of Management Review*, 23 (January), 14–31.
- Brown, BS. (1995) The academic ethics of graduate business students: A survey *Journal of Education for Business* 70, 151–156
- Brown, BS. (2000) The Academic Ethics of Graduate Business Students: 1993 to 1998 *Journal of Applied Business Research* Volume 16 no 4 retrieved April 26 2013 from <http://journals.cluteonline.com/index.php/JABR/issue/view/226>
- Brown, BS., Choong, P. (2005) An Investigation of Academic Dishonesty among Business Students at Public and Private United States Universities *International Journal of Management* Vol. 22 No. 2 June 2005 201
- Brown, BS., McNery, M. (2008) Changes in Academic Dishonesty among Business Students in the United States, 1999–2006. *International Journal of Management* Vol. 25 No. 4 December 2008 621
- Brown, BS., Weible, R. (2006) Changes in academic dishonesty among MIS majors between 1999 and 2004 *Journal of Computing in Higher Education* Volume 18, Issue 1, 116–134 <http://link.springer.com/article/10.1007/BF03032727>
- Campbell-Evans, G. & Leggett, BM. (2007) Plagiarism in a Cross Cultural Setting: Educating and Acting AARE Annual Conference. 27–30 November, Hobart
- Carroll, J. (2007) *I'm fed up with people treating plagiarism and cheating as the same thing*. Retrieved July 3, 2009 from http://www.brookes.ac.uk/aske/documents/Carroll_J_07_%20PerspectivePaper.pdf
- Carroll, J. & Appleton, J. (2001) *Plagiarism a good practice guide*. Oxford: Oxford Brookes University/JISC
- Clark, B. (2008) *History essay in the making*, interview by Jessica Shepherd, the Guardian 6th May 2008 Retrieved June 17 2012 from <http://www.guardian.co.uk/education/2008/may/06/highereducation.students>
- Darbyshire, P. & Burgess, S. (2006) Strategies for dealing with Plagiarism and the Web in Higher Education, *Journal of Business Systems, Governance and Ethics* 1:4, 27–41

- Dawson, MM. & Overfield, JA. (2006) Plagiarism: Do Students Know What it is? *Bioscience Education* Volume 8 retrieved on April 24 2013 from <http://www.bioscience.heacademy.ac.uk/journal/vol18/beej-8-1.aspx>
- Errey, L. (2002) Plagiarism: Something Fishy? Or just a Fish out of Water? *Teaching Forum* 50:17-20
- Erickson, BH. (1988), The Relational Basis of Attitudes, in *Social Structures: A Network Approach*, Barry Wellman and S. D. Berkowitz, eds., New York: Cambridge University Press, 99-121.
- Etter, S., Cramer, JJ. & Finn, S. (2002) Origins of Academic Dishonesty: Ethical Orientations and Personality Factors Associated with Attitudes about Cheating with Information Technology *Journal of Research on Technology in Education*, 39(2), 133-155
- Fallon, J. & Wellman, N. (2012) Investigating academic malpractice within an MBA Marketing Module *International Journal for Educational Integrity*, 8:1
- Flynn, S. (2003) Confronting academic dishonesty in the accounting classroom: a cultural experience? *Accounting Education* 12 (4), 437-439
- Gururajan, R. & Roberts, D. (2005) Attitudes towards plagiarism in Information Systems in Australian Universities. Retrieved April 26 2013 from http://eprints.usq.edu.au/856/1/Gururajan_Roberts.pdf
- Handa, N. & Power, C. (2005) Land and discover! A case study investigating the cultural context of plagiarism *Journal of University Teaching & Learning Practice* 2 (3) 64-84
- Hayes, N. & Introna, L. (2005) Cultural Values, Plagiarism and Fairness: When Plagiarism Gets in the Way of Learning *Ethics and Behaviour* 15 (3) 213-231
- Hayes, N., Introna, L. & Whitley, E. (2006) Plagiarism, values and computing: differing cultural expectations of academic work in information systems education. Retrieved April 26 2013 from <http://www.lums.lancs.ac.uk/files/sdaw/5707.pdf>
- Hayes, N., Whitley, EA. & Introna, LD. (2006) Power, knowledge and management information systems education: the case of the Indian learner. In *International Conference on Information Systems* (10-13 Dec 2006: Milwaukee, WI., USA).
- Iyer, R. & Eastman, JK. (2006) Academic Dishonesty: Are Business Students Different From Other College Students? *Journal of Education for Business* 82 (2) 101-110
- Johnston, B. (2003) The concept of plagiarism. HEA. Retrieved December 19, 2009 from http://www.heacademy.ac.uk/assets/York/documents/resources/resourcedatabase/id428_concept_of_plagiarism.pdf
- Jones, M. (2009) Back-translation: the latest form of plagiarism, 4th Asia pacific Conference on educational Integrity (4APCEI) September 2009 University of Wollongong NSW Australia. Retrieved January 1 2010 from <http://www.uow.edu.au/.../program%20of%20sessions%20and%20paper%20titles.doc>
- Lancaster, T. & Clarke, R. (2009) Contract cheating in UK Higher Education: promoting a proactive approach, Institutional Policies and Procedures for Managing Student Plagiarism, on 4th June 2009 at Oxford Brookes University. Retrieved December 31, 2009 http://www.brookes.ac.uk/aske/documents/Plagiarism09_LancasterClarke.pdf
- Lupton, RA., Chapman, KJ, Weiss, JE. (2000). A cross-national exploration of business students' attitudes, perceptions, and tendencies toward academic dishonesty. *Journal of Education for Business* 75(4): 231-235
- Magyar, A. (2009) Plagiarism Prevention: the dangers of separating management from pedagogy, Institutional Policies and Procedures for Managing Student Plagiarism on 4th June 2009 at Oxford Brookes University. Retrieved April 26 from <http://www.brookes.ac.uk/aske/plagiarism2009/>
- McCabe, ED., Dukerich, J. & Dutton, D. (1993) Values and Moral Dilemmas: A Cross-Cultural Comparison *Business Ethics Quarterly*, 3, (2). 951-960

- Megehee, CM. & Spake, DF. (2008) The impact of perceived peer behaviour, probable detection And punishment severity on student cheating behavior. *Marketing Education Review*, 18, (2) 5–19
- Miller, AT., Shoptaugh, C. & Parkerson, A. (2008). Underreporting of cheating in research using volunteer college students. *College Student Journal*, 42, 326–339.
- Moon, J. (1999), How to... stop students from cheating The Times Higher Education Supplement September 3, 1999
- Park, C. (2003) In Other people's Words: plagiarism by university students-literature and lessons. *Assessment and Evaluation in Higher Education*, 28 (5) 471–488. Retrieved on 22 April 2013 from http://www.lancs.ac.uk/staff/gyacp/caeh_28_5_021ores.pdf
- Schmitt, D. (2007) Academic writing ideas in flux, SDAW International Students, Academic writing and Plagiarism Conference September 2007 Retrieved April 21, 2013 from <http://www.sdaw.info/conference/abstracts/schmitt.ppt>
- Swift, CO. & Nonis, S. (1998) When no one is watching: cheating behavior on projects and assignments *Marketing Education Review*, 8, (1) 27–36
- Tennant, P. & Duggan, F. (2008) Academic misconduct benchmarking research project: part II: The recorded incidence of student plagiarism and the penalties applied. Amber Project Report Academy JISC Academic Integrity Service Retrieved 25 April 2013 from http://www.heacademy.ac.uk/assets/documents/AMBeR_PartII
- Timm, A. (2008a) Final report on educational practices in China, HEFCE: Student Diversity and Academic Writing Project. Retrieved 26 April 2013 from [w.info/lecturers/SDAW5China_Final%20Report%202008.pdf](http://www.sdaw.info/lecturers/SDAW5China_Final%20Report%202008.pdf)
- Timm, A. (2008b) Final report on educational practices in India, HEFCE: SADW Project. Retrieved December 17, 2009 from http://www.sdaw.info/lecturers/SDAW%20India_Final%20report%202008.pdf
- Underwood, J. & Szabo, A. (2003), Academic Offences and E-Learning: Individual Propensities in Cheating, *British Journal of Psychology*, 94 (August), 319–337.
- Wellman, N. & Fallon, J. (2010) Dealing with unacceptable writing strategies in an MBA marketing module retrieved 22 April 2013 from http://www.heacademy.ac.uk/resources/detail/internationalisation/Case_Story_unacceptable_writing
- Wilson, BA. (2008) Predicting Intended Unethical Behavior of Business Students. *Journal of Education for Business*, March/April 187–195
- Yeo, S. (2006) Science and engineering student's beliefs about plagiarism: It's only an assignment. UniServe Science Assessment Symposium Proceedings: Retrieved January 1 2010 from <http://science.uniserve.edu.au/pubs/procs/2006/yeo.pdf>
- Zey-Ferrell, M., Weaver, KM. & Ferrell, OC. (1979), Predicting Unethical Behavior among Marketing Practitioners, *Human Relations*, 32 (July), 557–569.

Appendix A: Summary data

Part C: Views on and participation in potential malpractice

		Agree					Disagree		%		% Others		OK OR NOT			
MEAN		very much	2	3	4	very much	n	I do	%	do	%	>3	%	>2	%	
c1	2.902	42	37	33	32	37	181	19	17.9%	19	17.9%	112	61.878	79	43.6	
c2	2.736	39	55	32	11	39	176	27	25.5%	27	25.5%	126	71.591	94	53.4	
c3	3.852	15	15	27	44	80	181	10	9.4%	33	31.1%	57	31.492	30	16.6	
c4	3.326	22	29	45	32	51	179	17	16.0%	22	20.8%	96	53.631	51	28.5	
c5	3.284	28	26	42	33	52	181	15	14.2%	29	27.4%	96	53.039	54	29.8	
c6	3.634	16	25	35	35	70	181	10	9.4%	25	23.6%	76	41.989	41	22.7	
c7	3.646	20	23	30	30	76	179	12	11.3%	28	26.4%	73	40.782	43	24	
c8	2.923	39	40	38	22	42	181	18	17.0%	22	20.8%	117	64.641	79	43.6	
c9	3.659	18	27	29	27	79	180	5	4.7%	28	26.4%	74	41.111	45	25	
c10	3.779	11	26	29	35	78	179	9	8.5%	26	24.5%	66	36.872	37	20.7	
c11	3.951	7	21	32	31	90	181	16	15.1%	26	24.5%	60	33.149	28	15.5	
c12	3.727	17	23	28	32	81	181	15	14.2%	23	21.7%	68	37.569	40	22.1	
c13	4.470	2	9	14	28	128	181	5	4.7%	26	24.5%	25	13.812	11	6.08	
c14	4.423	4	8	13	33	122	180	7	6.6%	31	29.2%	25	13.889	12	6.67	

QUESTION KEY: The practice is “OK”; I have done this; Others do this.

c1	It is OK if I don't know much about the subject of an individual assignment for others (students, seniors, friends, family) to help in any way they can
c2	It is OK if I was finding an individual assignment difficult, for others to help in any way they can
c3	It is OK if I was running out of time to submit an individual assignment for others to help in any way they can
c4	It is OK if I am asked by another student for help with their individual assignment, to help in any way I can
c5	It is OK for "seniors" and past UWIC MBA students to help current students with an individual assignment, in any way they can
c6	It is OK for people outside UWIC (friends, family students at other Unis etc) to help current students with an individual assignment, in any way they can
c7	It is OK to work on an individual assignment as a group as by sharing everybody stands a chance of getting a better mark
c8	It is OK to work on an individual assignment as a group as the stronger students can support the weaker ones
c9	It is OK to work on an individual assignment as a group as members can share the workload and thus save each other time
c10	It is OK if asked by other students, to work in a group and share material for an individual assignment
c11	It is OK to copy material from the internet then re-order it and/or change some of the words, as the changes make it your own work
c12	It is OK to copy material from a book or article then re-order it and/or change some of the words, as the changes make it your own work
c13	It is OK to copy material from a current student then re-order it and/or change some of the words, as the changes make it your own work
c14	It is OK to copy material from a past student, senior or friend then re-order it and/or change some of the words, as the changes make it your own work

Appendix B: Summary data

Part D: Participation and views on unfair practices

I have done this										Others do this		It is OK		Not at all		Perfectly				
	MEAN	Tot	Frqt	2	3	4	Rare	Never%	n			MEAN	Tot	OK	2	3	4	OK	n	
d1	4.719	807	12	9	21	13	39	77	45.0%	171	61	35.7%	2.095	287	59	30	30	9	8	137
d2	5.398	950	3	8	8	11	13	133	75.6%	176	52	29.5%	1.406	194	107	23	11	4	3	138
d3	5.361	965	3	3	10	11	36	117	65.0%	180	46	25.6%	1.604	223	93	23	18	6	6	139
d4	5.229	915	2	6	12	12	41	102	58.3%	175	47	26.9%	1.871	262	74	28	22	10	5	140
d5	5.164	914	10	7	13	9	18	120	67.8%	177	54	30.5%	1.555	213	96	16	17	3	4	137
d6	5.376	930	1	7	10	10	25	120	69.4%	173	46	26.6%	1.522	204	94	18	15	2	4	134
d7	4.836	856	11	12	20	9	25	100	56.5%	177	57	32.2%	1.696	234	85	23	19	5	5	138
d8	5.257	941	5	11	11	5	21	126	70.4%	179	57	31.8%	1.436	201	102	20	12	3	2	140
d9	5.182	912	8	7	15	9	13	124	70.5%	176	54	30.7%	1.714	240	90	18	18	6	7	140
d10	5.457	944	3	1	8	13	25	123	71.1%	173	73	42.2%	1.360	204	122	13	8	3	4	150

QUESTION KEY: I have done this; Others do this; It is OK to do this.

d1	Worked with others on an individual assignment
d2	Submitted work done by someone else as your own
d3	Asked to see the work of someone who has done a similar assignment
d4	Shown your own work to someone about to do a similar assignment
d5	Asked on a social networking site (eg Facebook) for help with an assignment
d6	"Padded" your text or bibliography with sources you've not actually consulted
d7	Directly copied material without referencing the author
d8	Submitted work taken wholly from an internet source (free or paid) as your own
d9	Translated sections of text to another language via web-translators (eg Babelfish)
d10	Taken credit for groupwork without doing a fair share of the work

Authors

Neil Wellman, Freelance education consultant, neil@nwellman.com

Dr Julia Fallon, Head of MBA, Cardiff Metropolitan University, Wales, jfallon@cardiffmet.ac.uk

Copyright © 2013 Authors listed on page 70: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

PAPERS—SECTION II
BEST PRACTICES AND STRATEGIES FOR AWARENESS,
PREVENTION, DETECTION OF ACADEMIC
MISCONDUCT

AN APPROACH TO DETECT ILLEGAL SIMILARITY IN RESEARCH LITERATURE USING LATENT SEMANTIC INDEXING

Muna Alsallal, Rahat Iqbal

Abstract: Research suggests that there are an increasing number of illegal similarities within research literature. As part of our research we are investigating the application of an information retrieval technique, Latent Semantic Indexing, to derive semantic information from text files. In this paper, we present an integrated framework for enhancing the automatic detection for illegal similarity texts, steering the area of Latent Semantic Index (LSI) and highlighting its ability to unmask the latent relationship between texts in order to detect illegal similarity. We have conducted an experiment to investigate the efficiency of a dimensionality reduction parameter as the core for LSI technique, the experiments designed to establish the highest amount of re-occurrence for given values and also the distribution for given values of the dimensionality reduction parameter k in Latent Semantic Indexing. The results so far are promising.

Introduction

The adopting of others' ideas and manipulation of the findings have overloaded research literature which makes it very difficult to optimize results of information retrieval and data mining systems to match user information needs. Some on-going studies suggest excuses for such a risky phenomenon stating that researchers may not have sufficient time to track their own ideas, and publishers may not be well-equipped to check whether the contributions and results come from original research (Alzahrany et al, 2012). Misconduct in research is becoming increasingly more sophisticated, including incidence of duplication of publications and illegal reuse of others' research ideas, contributions and findings, (Hennessey et al, 2012). There is scope for future research to address these problems by exploring use of algorithms to digitally analyse essential literature sources.

The Research Problem

The problem of illegal similarity in research literature is currently solved by a variety of different algorithms which function by comparing new manuscripts to the already published texts contained by a large library. This comparison works by means of comparing significant key words, phrases which are statistically improbable, and by computing a measure of similarity on the basis of aligning different sentences (Nayak, 2009), If this information exceeds a particular threshold, it may then alert the user to the possibility of illegal similarity.

However, there are a number of factors which need to be kept in consideration when deciding upon the effectiveness of an algorithm. For example, Lewis et al (2006) argues that the effectiveness of the algorithm will depend on some factors such as the number of databases which are served, the extent to which it is compatible with the journal

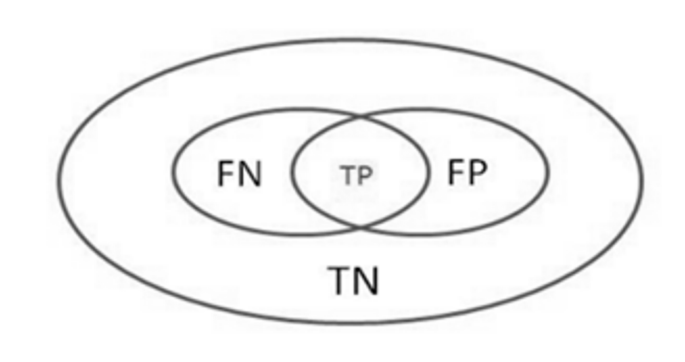


Figure 1. Venn diagram shows the current detection systems performance

manuscript submission system, the user interface (in particular, the extent to which the results are presented in a manner which is meaningful and easily assimilated) and the security of data. This is expanded upon by Garner(2011) who argue that the effectiveness of the algorithm at identifying illegal similarity is closely related to the specificity and the sensitivity of the search algorithm. It is therefore important to ascertain the way in which the search algorithm works, and what the false negative (documents that share suspicious text but not retrieved as suspicious) and the false positive (documents that is retrieved as suspicious but in point of fact it is innocent) rates of the algorithm are. Figure1 shows the current detection systems performance. Furthermore, for ease of use, it is suggested that an effective algorithm should provide the user with threshold settings in order to enable them to minimise false negatives and false positives and to prioritise different manuscript sections where it is possible to tolerate differing levels of similarity (Long, Errami & George, 2009).

Contribution

The contributions of our work are threefold: Section 2 identifies the concept of plagiarism detection systems and it's privacy within research environment. Section 3 simplifies the complicated concept of LSI which helps the readers from different backgrounds to understand the state of art of this pure mathematical method. In section 4 the proposed framework based on intelligent techniques and algorithms work together to detect the new added articles semantically and syntactically and define the old ones as well, in addition our framework has beneficiary of PPR (post publication review) technique which will give the user a chance to participate in enhancing the performance of the detection process, put this system of the interest of open access articles users and finally we are working to build a robust system to detect different kinds of illegal text reuse but keep consideration on preventing the loss of knowledge.

Illegal Similarity Detection Approaches

Recent anti-plagiarism systems can identify only word-for-word similarity and only some incidence of it (Eissen et.al, 2007) and do not cover the ideas adoption or fake results. On-going research in this area classifies plagiarism detection systems into two types:

- Extrinsic plagiarism detection systems which deal with declared illegal text similarity as most of algorithms work using a technique to compare word for word to discover illegal similarities syntactically and some algorithms attempt to detect illegal similarity semantically. However these systems are gaining good results such as Turnitin which is used widely as a detection system for student essays, but this tool still faces difficulties in detecting some kinds of text manipulation. Approaches used to detect extrinsic illegal similarity required a reference collection of likely original texts.
- Intrinsic plagiarism detection systems work to unmask undeclared illegal text similarity, such as ideas adoption or fake results. In intrinsic detection approach, researchers attempt to simulate the human judgment in discovering the intelligent changes in some texts although they do not have a reference library in their memory. Stylometric feature extraction is one of the interesting research approaches that will be used in our proposed framework to enhance the detection process.
- English language becomes the crucial language for publishing stakeholders, that makes reuse some words or phrases in order improve the language and writing style a common practice for international researchers to deliver their ideas and findings in an acceptable style. However most current plagiarism detection identify the above practice as an illegal despite the originality of ideas and findings (Mason, 2009).

Latent Semantic Indexing (LSI)

LSI is mainly used as an information retrieval technique, which is generally based on the spectral analysis of the term document matrix (Edmunds 1997). It essentially captures the hidden structure within an information retrieval method using techniques from linear algebra (Veling and Van der Weerd 1999). What generally occurs is that there are vectors that represent the documents or text, which are projected in a low dimensional space which is then obtained by a singular value decomposition of the terms (Dhillon and Modha, 2001). The most interesting feature of LSI is language independent

How LSI in our Proposed Framework Works

The procedure has been carried out by implementing five steps:

1. Pre-processing by eliminating sole and popular terms from the corpus and activate stop word list which contains conjunctions, prepositions..etc
2. The functionality of LSI starts by transforming the pre-processed corpus into an $m \times n$ matrix in which its column represent documents and its rows represent terms (Berry et. al, 1993)

3. Apply term weighting scheme to eliminate the non-relevant terms as local weighting shows the term frequency in document while global weighting shows the term frequency in the corpus.
4. Apply SVD method to reduce the corpus into an intensive corpus containing only significant terms from the original by decomposing the original matrix into 3 concentrated matrices (Berry et. al, 2005).
5. Apply Dimensionality Reduction technique in order to truncate the three matrices that resulted from step 4 into k-dimensions by selecting the first k columns from each matrix and assign zero for the rest values.

Effective Parameters

Previous manuscripts have shown promising results while using latent semantic indexing to simulate the human ability to detect illegal similarity. The methods that have been conducted using LSI are quite different and resulted in putting huge obstacles to identify which parameters will lead to success or failure (Haley et al., 2005, 2007). We can summarize the most important parameters that may affect the performance of LSI into four groups:

1. Weighting schemes, which represent the importance of the term in document that so-called local term weight and global term weight which represents the importance of a specific term in the corpus. Wild et al. (2005) advised to use the formula (1) to calculate the local term weighting as they had achieved good results.

$$\text{Log} \quad l_{ij} = \log(tf_{ij} + 1) \quad (1)$$

where tf_{ij} is the number of times that term i appears in document j .

For global weights, formula (2) is the common global weighting formula has been used. However (Harley et al. 2005) had achieved a very satisfactory output by using formula (3)

$$\text{IDF} \quad g_i = \log_2(n/df_i) + 1 \quad (2)$$

where df_i is the number of documents where the term i appears

$$\text{Entropy} \quad g_i = 1 + j(p_{ij} \log(p_{ij}) / \log(n)) \quad (3)$$

where tf_{ij} is the number of times that term _{i} appears in document _{j} ; gf_i is the total number of times term i appears in collection; n is the number of documents in the collection.

2. Dimensionality Reduction: Many researchers elucidate that dimensionality reduction represent the mathematical core for LSI so it is valuable to investigate this parameter to establish the usefulness of LSI method. Since the idea of LSI depended on establishing the term-document matrix to represent the text and using SVD; the mathematical method to truncate the original matrix into k most relevant dimensions in order to eliminate the non-relevant documents (Zeimpekis

& Gallopoulos, 2005). Far from mathematical language, DR reduces large datasets to intense datasets containing only related data from the original data.

3. **Similarity Measures:** Many measure correlations have been used to measure the distance between two vectors, for instance the common similarity measure is cosine measure which practically widespread for plagiarism detection. The automatic detection systems are using the cosine measure between the suspicious document vector and relevant documents.

$$\cos(VS1; VS2) = \sum_{i=1}^k (VS1_i \cdot VS2_i) \frac{VS1_i * VS2_i}{|VS1| * |VS2|} \quad (4)$$

where $VS1$ is the vector representing the first suspicious file, $VS2$ is the vector representing the second suspicious file and k is the number of dimensions.

4. **Pseudo documents:** Pseudo documents are constructed to represent the content of documents in semantic spaces in order to extract the underlying relationship between terms. We can construct pseudo documents using two popular methods:
 - **Vector Sum:** Extract the meaning of documents via summation of word vectors of the document.
 - **Folding-in procedure:** Constructing the pseudo document via folding an extra document into k -dimensional vector space.

Outline of Framework

We propose a framework in order to increase the effectiveness in which illegal similarity in research literature can be identified. The proposed framework consists of group of components working together to enhance the detection task.

The framework which is outlined above is a proposal for automatic illegal similarity detection (ISDT) which fulfills the task of detecting illegal similarity semantically to declare the latent relationships between terms. It consists of the following components:

- **Latent semantic indexing (LSI)** which fulfills the role of detecting the incidence of semantic similarity between different texts.
- **Modified Stylometry algorithm** which fulfills the role of detecting the similarity between different texts depending on extracting the writing style features.
- **Modulator** which works by calculating the extent of similarity between two separate texts based on a given threshold value. Once this rating has been calculated, this information then entered onto the global similarity index.
- **Global similarity index.** It works as a management component that stocks the suspicious detected files in the master list of detected
- **Medline index** which is the principal index pertaining to the research literature
- **A keyword based search tool** which works by retrieving all of the documents which are relevant to a particular query.
- **Normal User interface** which consists of a search system interface which is employed for the searching of documents by users.

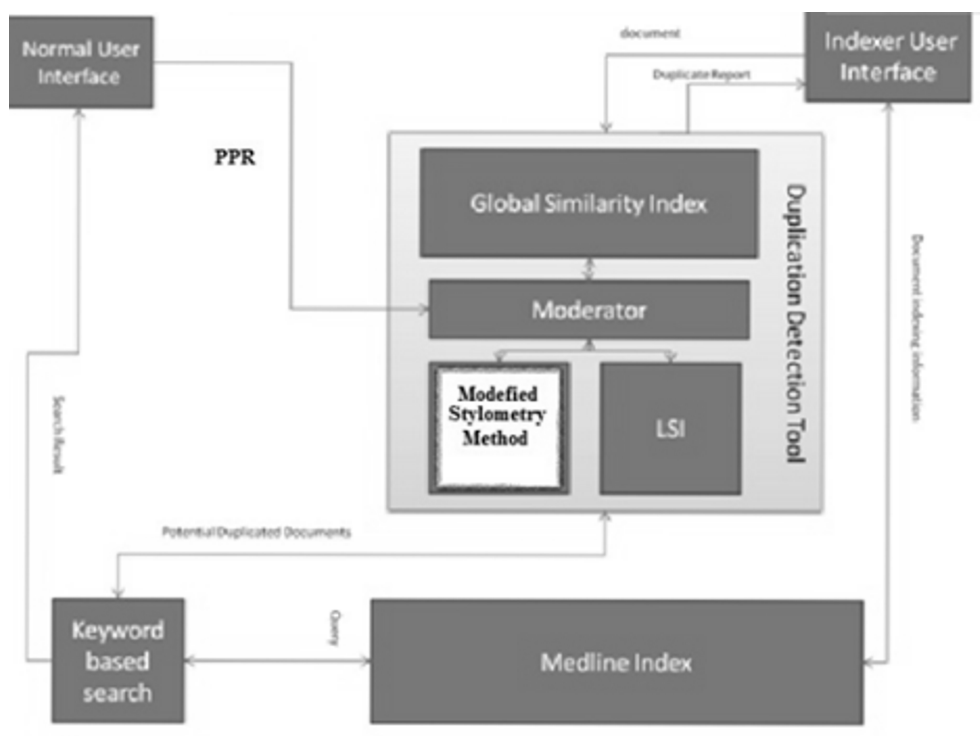


Figure 2. The Proposed Framework

- Post Publication Review (PPR) technique which will be used to get the user feedback on the results from ISDT. This opinion will be weighted and entered into moderator calculations.
- Indexer user interface which is used by individuals who add information to the Medline Index.
- PPR Technique: the user receives the results of the search together with a list of associated optional contextually alike documents. At this point, the user will be asked to manually assess the documents which have been retrieved to determine the potential identical documents in order to ascertain whether the level of similarity between both documents is sufficiently high; this manual assessment so-called post publication review technique (PPR) can be used as an important factor to enhance the detection process. PPR factor will in-turn be employed in moderator calculations. The identical similarity process will be iterated taking into its account the value of PPR factor.

Method

According to Lewis et al (2006), the key distinguishes features of LSI is the fact that the meaning of terms is derived by means of an approximation of the structure of term

Table 1
Deerwester (Deerwester et. al, 1990) co-occurrence / Illegal similarity summary

Degree	Number of Occurrences / Duplicates
1	30
2	22
3	12
4	3

usage in different documents using the aforementioned decomposition algorithm. We conducted two stages experiments the first stage without using LSI method and the second stage with using LSI method.

Experiment Details

The experiments were carried out to investigate the significance of using dimensionality reduction parameter to highlight the bridge terms and its importance in unmasking the hidden relationships between two terms that don't coincide together but instead they coincide with a joint term so-called a bridge term. We conducted the experiments using MED collection for both of experiments to investigate the performance of LSI comparing with Find-Transitivity program.

MED Collection

The collection has been used to find singular values and vectors for $k = 50, 100, 200$ and 300 . The singular values and vectors were then be an input for find-transitivity program; this identifies the level of transitivity for each pair of identical pairs (Swanson, 1991). Further work was carried out which developed a term by term duplicate matrix. The find-transitivity program takes an input in the form of vectors and singular values produced by the SVD program.

Results and Discussion

A summary of the results found by using the find-transitivity program is shown as follows:

From stage (1), four was the highest order that was observed in the case of the find-transitivity program for the MED data.

From stage (2) the following table shows the results up to third order co-occurrence, which was examined.

The table above shows the results for the MED inputs to degree level 3. There were no significant changes in term of k , however it was noted that there was a slight upwardly trend, when k increases. The MED data reveals that there are a total of 1,110,491 pairs of terms that can be accessed or connected directly. A total of 15,869,045 pairs of terms are connected when one intermediate node is allowed, whilst there are 17,829 pairs of terms that can be connected with two intermediate nodes. Each of the terms within

Table 2

Results from MED Testing

	Degree 1	Degree 2	Degree 3
k = 50	1,110,483	15,866,518	17,817
k = 100	1,110,485	15,867,200	17,819
k = 200	1,110,491	15,867,595	17,824
k = 300	1,110,491	15,867,595	17,824

the MED data can be connected to one another with two intermediate nodes or less between terms.

It is clear from the evidence shown that transitivity within the co-occurrence relationship plays a key role in ensuring the effectiveness of any system that uses tools such as information retrieval, computational linguistics and textual data applications. Results from both types of experiments carried out have concluded that there is a means of improving the illegal similarity detection within a system in a successful manner, there is grounds for further investigations to combine the Dimensionality Reduction (DR) parameter to enhance the results as DR perform well when connects terms through bridge terms nevertheless in term of k different values doesn't give a motivating pointer. In addition the optimum of parameters combination will increase the effectiveness of LSI method to detect illegal similarity between research papers. Further experiments using different parameters will be carried out in order to achieve the goal of the proposed framework.

References

- Berry, Michael, Theresa Do, Gavin O'Brien, Vijay Krishna, and Sowmini Varadhan. 1993. SVDPACKC (Version 1.0) User's Guide. April 1993.
- Berry, MW. Pulatova, SA. & Steward, GW. 2005. Computing sparse reduced-rank approximations to sparse matrices, *ACM TOMS* 31, no. 2.
- Cosma, G. & Joy, M., 2006. Source code plagiarism: a UK academic perspective, *Research Report*, No. 422, Dept of Computer Science, Univ of Warwick, Coventry
- Cosma, G., 2008. An approach to source code plagiarism detection and investigation using latent semantic analysis, Thesis, University of Warwick, available from <http://www.dcs.warwick.ac.uk/report/pdfs/cs-rr-440.pdf> accessed 30 January 2013
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6): 391-407.
- Dhillon, IS. & Modha, DM. 2001. "Concept Decompositions for Large Sparse Text Data using Clustering", *Machine Learning*, 42:1, pp. 143-175.
- Edmonds, Philip. 1997. Choosing the Word Most Typical in Context Using a Lexical Co-occurrence Network. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics. Pages 507-509.
- Errami, M. & Garner, H. R., 2008. A tale of two citations, *Nature*, vol. 452, pp. 397-9

- Garner, H. R., 2011. Combating unethical publications with plagiarism detection services, *Urologic Oncology: Seminars and Original Investigations*, vol. 29, pp. 95–99
- Gollogly, L. & Momen, H., 2006. Ethical dilemmas in scientific publication: pitfalls and solutions for editors, *Rev Saude Publica*, vol. 40, pp. 24–9
- Haley, D. T., Thomas, P., Petre, P., & De Roeck, A. (2007). Seeing the whole picture: Comparing computer assisted assessment systems using LSA-based systems as an example. Technical Report Number 2007/07.
- Hennessey, K. Williams, A. Afshar, A. MacNeily, A. (2012) Duplicate publications: A sample of redundancy. *Journal of Urology*, 6(3), pp. 177–180
- T. C. Hoad and J. Zobel. Methods for Identifying Versioned and Plagiarised Documents. *JASIST*, 54(3): 203–215, 2003.
- Lewis, J., Ossowski, S. & Hicks, J., 2006. Text similarity: an alternative way to search MEDLINE, *Bioinformatics*, vol. 22, pp. 2298–304
- Long, T. C., Errami, E. & George, A. C., 2009. Scientific integrity: responding to possible plagiarism, *Science*, vol. 323, pp. 1293–4
- Mason PR. Plagiarism in scientific publications. *J Infect Dev Ctries*. 2009;3:1–4.
- Meyer zu Eissen, S., Stein, B., & Kulig, M. (2007). Plagiarism detection without reference collections. *Advances in Data Analysis* (pp. 359–366) Springer.
- Moussiades, L. M. & Vakali, A., 2005. PDetect: A clustering approach for detecting plagiarism in source code data sets, *The Computer Journal*, vol. 48, pp. 651–661.
- Mozgovoy, M., 2007. Enhancing computer-aided plagiarism detection, Dissertation, Department of Computer Science, Univ of Joensuu.
- Nayak, B. K., 2009. Author's misconduct inviting risk: duplicate publication, *Indian Journal of Ophthalmology*, Nov–Dec, 57(6), pp. 417–418.
- Schütze, Hinrich. 1998. Automatic Word Sense Disambiguation. *Computational Linguistics*, Volume 24, number 1.
- Swanson, DR. 1991. Complementary Structures in disjoint science literatures. In A. Bookstein, et al (Eds), *SIGIR91: Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*. pp 280–289.
- Veling, Anne and Peter van der Weerd. 1999. Conceptual grouping in word co-occurrence networks. *Proceedings of the IJCAI '99*. Volume 2. Pages 694–699.
- Zeimpekis, D. & Gallopoulos, E., 2005. TMG: A MATLAB Toolbox for generating term-document matrices from text collections, in *Grouping multidimensional data: recent advances in clustering*, J Kogan & C. Nicholas ed. New York, USA: Springer Publishers.

Acknowledgements

Personally, I would like to thank Mrs Irene Glendinning; my line manager for encouraging me all the time in addition to her continues support. Without her care; was not easy to overcome hard times. I truly appreciated all her time and advices.

I would like to thank Dr. Rahat Iqbal and appreciated his guidance, support and willingness to take time to discuss my research.

Authors

Muna Alsallal, aa6095@coventry.ac.uk, Rahat Iqbal r.iqbal@coventry.ac.uk, Coventry University, United Kingdom

About authors

Muna Alsallal—BEng in Computer Hardware and Architecture Msc in IT and Application

PhD Candidate in Computing and Engineering Coventry University.

Dr Rahat Iqbal is a Reader in Human-Centred Technology (HCT) and Director of Applied Computing Research Centre (ACRC) in the department of computing at Coventry University.

His research interests lie in information retrieval and supportive systems, in particular with regard to human factors and personalization of these systems based on user needs and context. Currently, he is the principal investigator of two funded projects including the EPSRC funded project tilted task sensitive search recommender system based on implicit feedback using intelligent techniques and a Technology Strategy Board funded project on driver's behavior with Jaguar Land Rover Ltd. He has published more than 70 papers in peer reviewed journals and reputable conferences and workshops. Dr. Iqbal is on the programme committee of several international conferences and workshops and organises the annual international workshop on Ubiquitous and Collaborative computing in conjunction with the BCS Conference on Human Computer Interaction. He is also a panelist and fellow of the UK Higher Education Academy and a member of the Nuffield Foundation (Science Bursaries & Oliver Bird Rheumatism Programme). Dr. Iqbal has also been editing several special issues of international journals within the field of information retrieval and user supportive systems.

Copyright © 2013 Authors listed on page 93: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

UNDERSTANDING FIRST YEAR UNDERGRADUATE STUDENT PERCEPTIONS OF COPYING AND PLAGIARISM: DEVELOPING A PLATFORM FOR A CULTURE OF HONEST INQUIRY AND THE ACADEMIC CONSTRUCTION OF KNOWLEDGE

Robert Craig, David Dalton

Abstract: Academic integrity is a concern in tertiary institutes globally with some perceiving it as endemic. Amongst the issues raised by this practice are serious ethical considerations. Impacts on learning are also evident in that if students are not doing their own work, what is actually being learned? The reasons why students plagiarize are therefore a legitimate area of study and may not be as straightforward as they first appear. The main objective of our research was to gather data which would highlight the perceptions first year students have concerning the proportion and frequency of cheating among their peers. There was also interest in discovering what students considered to be the main reasons for such behaviour. Findings show that at the start of the undergraduate programme, three quarters to four fifths of the students viewed copying as serious or very serious. However, after only one semester this percentage had dropped considerably for some areas. Frequency of cheating also varied from the first to the second semester. Such data has resulted in considerations of what kind of changes might be required in order to minimize the perceived “need” to copy or plagiarize, and generate institutional discussion on the issues. The paper highlights proposals for structural and content changes to curriculum focus, delivery, and the learning environment, based on current good practice in these contexts within our institute, in particular, concepts and practices of student ownership and buy in which we believe removes a significant part of the “need” for copying and plagiarism.

Introduction

The study described was carried out in response to expressions of concern about copying and plagiarism, similar to those encountered in many tertiary institutes worldwide. That copying and plagiarism are a major problem in many spheres would not be contested. Headline cases have included the suspension of Fareed Zakaria, the case of Romanian Prime Minister Victor Ponta, British journalist John Hari and locally, in our context, Dr Hafnaoui Ba’li who lost the prestigious Sheikh Zayed book award as a result of plagiarism. In higher education specifically it is considered by many to be “endemic in universities worldwide” (Sheard et al, 2002, p183) and not only amongst students, with the University World News for example, claiming that research in Egyptian universities is “plagued by plagiarism” (Khaled, 2008, p1). This statement could probably be reasonably generalized to other parts of the world.

Our concern as educators and researchers is to focus on the behaviours of undergraduate students in respect to these practices and explore perceptions to copying and plagiarism amongst our student body. Recent research into plagiarism and copying in universities supports anecdotal perceptions that it is very widespread and varied in

nature. A survey amongst 150 undergraduates conducted at the University of Pretoria in 2005 found that 80% of the respondents admitted to “frequently” plagiarizing assignments (Russouw, 2005). According to one British newspaper (Daily Telegraph, 2012), data derived from the university admissions system (UCAS) in the UK in 2011 revealed 8,500 students even copied personal statements written in their university applications! A survey conducted at Monash University in 2001 found that 85% of respondents admitted to cheating with 33% stating that they copied from friends’ assignments while doing some of the work themselves, 10.4% copied assignments from another students’ computer without their knowledge and 22.4% copied from the internet (Sheard, 2002).

Donald McCabe is perhaps one of the leading researchers in this area, certainly in the US and Canada, and has conducted research with samples of up to 71,000 undergraduate students over a decade (McCabe, 2005). He has found that “cheating” of various kinds is widespread and habitual. He has looked at a range of different academic environments where copying and plagiarism occur, including examinations, written assignments, research work and lab work. He found that, 42% of respondents worked with others on an assignment when asked for individual work, 38% paraphrased or copied from a written source without footnoting and 36% did likewise when using the internet. Other elements were handing in assignments copied from other students and falsifying bibliographies. Exams provided another fertile area for suspect behaviour with 11% copying from another student during an exam, 8% using unauthorized notes and 5% using unauthorized digital or electronic devices. Recently the prestigious Harvard University recently launched an investigation after “possible” cheating was discovered in roughly half the papers of a class of 279 students in the spring semester of 2011. The picture painted from the above cited research and supported by other studies as well as the day-to day discussions that take place between educators everywhere clearly shows patterns of behaviour that are universal and apparently habitual.

Given that the practice is so common and gives rise to a wide range of concerns, it is important to identify and discuss some of the issues and dilemmas it raises. These include ethical, philosophical, cultural, social, and pedagogical questions as well as others. Academic integrity, which most institutes of higher education would at least claim to have as one of their “core values”, requires honesty and fair play in the production of original documentation and in general academic practice. Copying the work of others whether in an examination or from the internet or other sources, breaches ethical guidelines and considerations and is reasonably considered to be a form of theft. Turning in plagiarized or copied work is also arguably unfair to those who produce their own work given that the “copier” may receive a higher grade for work they did not do. This in turn undermines the integrity of the grading process and assessment system given that assessment in large part evaluates what an individual has (supposedly) learned.

Another issue is the effect on learning itself, which of course is the reason students attend university and the reason teacher teaches. Plagiarism and copying of any kind, and particularly when it becomes habitual, means that valuable skills which form the core of the pedagogical process may well not be learned. For example, the research process involves a host of skills essential to that process but also skills for life. Evaluating

sources, reading and note taking, annotating texts, paraphrasing and summarizing are all skills which might be used later in a professional (work) environment. The process also involves key critical thinking skills especially in a discussion and recommendations section of a research report. If students are plagiarizing and these skills are not effectively learned there is also an obvious effect on the quality of their preparation for later endeavours including post graduate study and work. Apart from possible “knowledge gaps”, ethical issues are also raised, in that if students are little concerned about academic integrity, what might this suggest about pre-disposition to unethical behaviour in a professional environment? A study conducted by the Josephson Institute (2009) looked at the relationship between attitudes and behaviour in high school and later conduct in adult life. It found that those who cheated on exams in high school twice or more are considerably more likely to be dishonest in later life, compared to those who do not.

More immediate consequences for the student who is caught usually involve a range of penalties increasing in seriousness which can ultimately result in expulsion from their university. This is as well as the stigma of being publicly branded a “cheat” which would likely appear on the student’s personal record and could obviously impact on their future life.

For those involved in a dynamic of copying and plagiarism, there are obviously reasons that generate such behaviours and personal perceptions related to these practices. It is likely that some forms of cheating will be viewed more seriously than others and indeed that some of the behaviours described may not be seen as cheating at all.

These were some of the issues focused on in our investigation at the Petroleum Institute (PI) in Abu Dhabi, the capital of the United Arab Emirates. At present, baccalaureate and post-graduate degrees are offered in Chemical Engineering, Electrical Engineering, Mechanical Engineering, Petroleum Engineering, and Petroleum Geosciences Engineering. Set up in 2001, it is in the unique and fortunate position of being supported, financed and governed by a national oil company and its international partners. Approximately seventy per cent of the undergraduates are Emirati and thirty per cent expatriate. While the degree programmes are open to both male and female students the two are, for the most part, segregated in order to comply with cultural sensitivities. The medium of instruction is English, (as it is throughout higher education in the UAE) and for most of the student population it is an additional language. While some may be quite fluent, many have a relatively low level of language proficiency on entry to the degree programme. It is reasonably fair to say that high school does not prepare them particularly well for the rigors of undergraduate study. The investigation described below was carried out in response to expressions of concern about copying and plagiarism, similar to those encountered in many tertiary institutes worldwide.

Methodology

The main objective of the research was to gather data which would highlight the perceptions first year students at the PI have concerning the proportion and frequency

of cheating among their peers. There was also interest in discovering what students considered to be the main reasons of such behaviour. A survey comprising five sections was developed to gather quantitative data which would answer the following four questions:

- Is academic dishonesty an issue at our institution?
- How serious do students consider it?
- How frequently do students believe it occurs?
- What are the reasons for such behaviour?

A seventy-five per cent response rate from a total of 185 first year students was recorded, comprising 87.5% of first-semester male students, 80% second-semester female students, and 57.5% of second-semester males. As such we were able to consider gender differences as well as any changes in students' perceptions of behaviour that may have occurred between entering the degree programme and after the beginning of the second semester.

Main findings

1. The first section of the survey was concerned with finding out if students were aware of the institute's academic integrity policy and honour pledge. The vast majority (90%), of both second-semester male and female students stated they were, whereas just under a third of first semester students claimed that they were not. Only half of these students had actually read the document. Four times as many male as female students surveyed said they did not actually understand it.
2. In response to the focus of the second section on student perceptions of the seriousness of academic cheating it can be surmised that approximately three quarters to four fifths of students who had recently begun their undergraduate programme believe that the items for consideration—copying homework, in a test, a colleagues answer, providing answers or doing homework for a friend or colleague, as well as plagiarizing work and passing it off as their own and paying someone to do their work—are serious or quite serious academic offences. It was, however, surprising to discover that after one semester, this percentage had dropped considerably in some areas. For example, half of male and female students did not consider copying homework, doing the homework of a friend or having the friend do it for them to be serious. Any form of cheating in exams, quizzes or tests was still perceived to be a serious academic offence.
3. The data also showed a change in the frequency of cheating from the first semester to the second. Almost all first-semester students claimed that they rarely or never copied homework from a colleague or from notes in a test nor did they provided answers for their friends or pay someone to do their work. Just over a month into their second semester, around half of both male and female students stated that they frequently or sometimes copied homework assignments, and over a third of males allowed friends to copy homework and quizzes.

4. There is a strong belief, among second-semester male and female students in particular, that cheating in various forms is quite rampant, especially on homework assignments.
5. The final section of the survey asked students to identify what they considered to be the most likely reasons for poor academic integrity. The choices included poor time management, a low chance of being caught or reported, minimal penalties, worth the risk to get a better grade, normal behaviour, having actually been taught to copy in high school, lack of understanding of how to complete the assignment, too much effort required, the author's words are best, poor command of English, and a general lack of interest. For first semester students the most likely reasons were a lack of proper understanding of how to complete the assignment and it being worth the risk to get a better grade. This was followed by a general lack of interest, a lack of time, low penalties and the belief that the author's words were in fact too good to change. A low chance of being caught, too much effort required to paraphrase or their English not being good enough, did not rate as highly as probable reasons for cheating.
6. A semester of undergraduate study showed some changes in most probable reasons for such behaviour. Poor time management was clearly the most likely reason for students to cheat. Students were aware of the penalties but considered it worth the risk to achieve a higher grade. This was followed by a poor understanding of the task requirements.
7. While gender differences were generally not extreme, varying no more than five to ten per cent on most responses, data does suggest that females take cheating more seriously. For example, twice as many females as males stated that they had actually read the institute policy on academic honesty, and ninety per cent of them believed that they fully understood it as opposed to about two thirds of the male students. Ten per cent more females than males consider cheating on exams, quizzes or tests, and helping friends in them, to be serious or quite serious, and two thirds of the female respondents and only half of the males believe that cheating on homework is serious.

Discussion

The main findings described above are consistent with other studies (Bjorklund & Wenestam, 1999 and Akande 1998), with poor time management and low understanding of the task ranking high. One might assume that students studying in a second or additional language often struggle with tasks and this may cause some temptation to cheat but our results show that "lack of understanding of how to complete the assignment, too much effort required, the author's words are best, and a poor command of English" were not rated highly by respondents as causes for cheating. This may well be due to the nature of the curriculum in their chosen majors in engineering with a heavy focus on traditional math, chemistry and physics. However, language may be an additional factor here as anecdotal evidence indicates misunderstanding based on ineffective decoding of teacher language (one of the researchers is currently gathering

qualitative data on this area) with an additional cultural element of students often not asking for clarification as they do not want to lose “face”.

It is not surprising that students do not read the policy—length and sophistication level of the document, and the fact that they have rarely had opportunity to engage in critical reading is hardly conducive to satisfactory compliance. Knowledge of policy does not, however, seem to be a factor in modifying behaviour with respect to cheating and plagiarism, perhaps in the same way that knowledge of traffic policy often has no impact on driver behaviour. Sheard et al also indicated high levels of student awareness of policy (84%) but little effect of this on levels and patterns of cheating. A further issue in our environment was the high number of students who did not understand the policy anyway. A need for more accessible language given that our students are native Arabic speakers may be required here as the style and lexis of “policy language” can sometimes be rather mystifying, even to native speakers of the language the document is expressed in.

A low chance of being caught ranked fourth in the above study, but was not ranked highly for our sample. In the context of written assignments, particularly in required first year Communication courses at the PI, plagiarism detection software is used widely in some course but not at all in others. This may explain the student perception that chances of being caught are low. Of course, it may also be that faculty might not be vigilant in this area. The findings do suggest that students understand the moral aspects of cheating but are selective in proportioning seriousness. Frequency changes in cheating between semesters can in part be explained by students prioritizing which elements of which course to focus on in order to achieve a desired grade. This is not a perspective students arrive with in the first semester, but seems to be learned behaviour. By the second semester they have learned to develop strategies (possibly influenced by peers) to achieve a “percentage” outcome. Testing appears to be taken far more seriously than learning and students develop strategies to deal with pressure and what they see as low priority. Unfortunately, homework and engaging in the learning process does not appear to merit the same worth as cramming for quizzes, tests and exams and this is likely to be the result of the culture of high schools in the region. It also perhaps due to the fact that many students at the beginning of the transition from high school to university do not understand that one of the main purposes of higher education is to foster intellectual growth through disciplinary inquiry; they tend to understand their new environment in terms of their somewhat limited high school experience of passing courses through the memorization of knowledge, tested through quizzes and exams. This surface level approach to learning is both intellectually and physically less demanding, and if it is tolerated will likely provide far more opportunity for cheating than a curriculum which encourages deep learning (Akande, 1998). Recent moves towards a more active, experiential and inquiry-based paradigm may well foster deeper learning, responsibility for learning and greater motivation, as opposed to the more traditional approaches which tend to encourage an the rewards of depending on surface learning strategies.

One thing that did surprise us a little was that students who generally (if you ask them) would describe themselves as religious, even defining themselves as such before other labels, did not seem to have any behaviour or perceptions of cheating as unethical

that were any different from studies carried out with students in more (nominally) secular societies such as the US. Rettinger's (2005) study indicates that

"among these religious students, more religiosity correlates with reduced reports of cheating in all courses. This result appears to be due to the unique effect of religion on self-reported cheating rates and, depending on course content, on a reduction of grade orientation in religious students" (p107).

We might have reasonably expected to find something similar here, but it was not apparent in our results. Other factors such as Hofstede's (2001) uncertainty avoidance (if accurate) might also lead us to assume that cheating in general might be less likely to occur, especially given (under the social norms trait) a supposed tendency towards conservatism and a belief in law and order. Ergo, if your institute has a stated policy (law) against plagiarism and cheating you should follow it. This does not seem to be the case here, even though most local universities have academic integrity/honesty policies. For example, a local study (Gulf news, 2012) indicated that 78% of respondents admitted to some form of cheating.

In general terms, the relationship between globalism and secularism is well explored (Beyer, 1994 and Alvey, 2003) and while it is not the purpose of this paper to enter the debate, secularization may well play a part in why our students behave in the same way as students anywhere. Further study would reveal the significance of this affective factor. Indeed, Hatherley-Greene (2012) states that in the UAE, we might be "witnessing... a neo-indigenous effect produced by globalization and the UAE's own cultural tsunami."

Research has also posited several possible other indicators of the propensity to cheat. One of these is the theory that those who come from collectivist societies, societies which are considered to manifest high uncertainty avoidance, and with a synchronic/short-term perception of time are perhaps more likely to adopt surface level approaches to learning, focusing on correct answers rather than process and analysis—a more individual approach favoured by the west (Hofstede, 2001). However, Hatherley-Greene (2012) believes that undergraduate students in the UAE do not consistently conform to the stereotype of the typical Arab cultural pattern yet still demonstrate quite a high level of cheating in areas they do not rate seriously, such as copying homework.

While ethical considerations and approaches to the "hearts and minds" aspect of this problem should certainly be sustained, there are clearly issues with the usefulness of a uniquely ethics-based approach. Similarly a lot of attention is given to detection and punishment. It may well be that both of these somewhat miss the point and that answers lie more in the sphere of curriculum, teacher /student relationships, classroom organization /management and the very nature of the courses and learning experiences we offer our students.

Approaches to creating a culture of academic integrity

How, then, do we change perceptions of learning and how do we move towards a more critical, deep learning approach when rote learning and the belief that there is one correct answer is entrenched? Much of the design of the two first year Communications

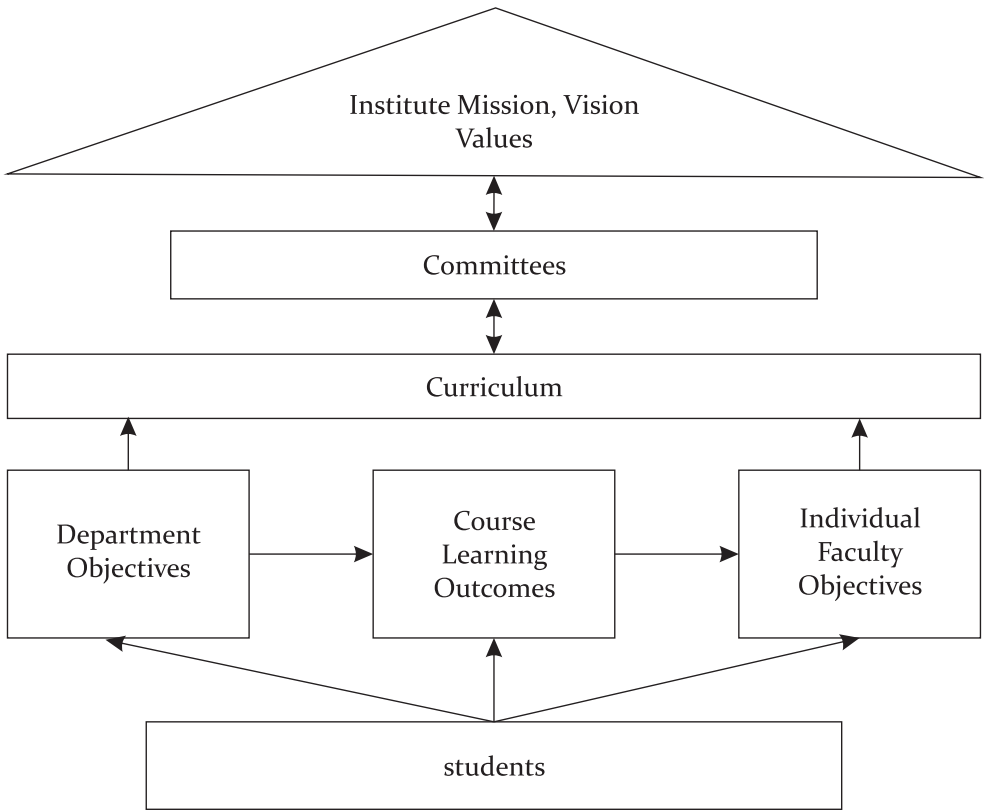


Figure 1. Responsibility for academic integrity

has been informed by two sets of learning principles—McCabe and Pavela’s (2007) ten principles of academic integrity, and Chickering and Gamson’s (1987) principles of good educational practice. As shown in Fig.1 below, academic integrity is the duty and concern of all members of the community.

One part of the PI’s mission statement (Petroleum Institute, Abu Dhabi, 2003) is to foster an “intellectual environment that leads to the development of our graduates as whole persons and as the future leaders in their respective fields of expertise.” One aspect which reflects the whole person development of the graduating student is that they “should have the professional integrity and maturity to serve humanity and its highest values, and should always make ethical decisions as they relate to society, corporate operations, technology, and the environment.” An academic honesty committee exists with the duty of considering incidences of cheating, recommending suitable sanctions, and more recently to make recommendations based on data for improvements to policy. A code of honour also exists, and students are required to sign a pledge attesting that each submission is their own work, and serves as a reminder of the institute’s commitment to maintaining a culture of academic integrity.

Documentation regarding widespread cheating in higher education however, suggests that policy and sanctions alone do little to change student attitudes to copying and plagiarizing. When the focus of much current practice is still very much on the ability to recall content, and provide answers rather than describe processes, there is abundant opportunity for students to resort to copying. Perhaps a move towards a more active, experiential and inquiry-based paradigm is more appropriate approach if learners are to become more responsible for their own learning, explaining understanding, and developing critical thinking.

The first year Communication courses have been developed to compensate for the poor preparation many of our students have for the requirements of twenty-first century undergraduate study. While the main objectives are to improve communication skills, enabling objectives include developing appropriate study skills, and applying critical thinking to reading and writing. The general philosophy is that cognitive academic language proficiency best develops through the acquisition and articulation of knowledge and understanding. The courses introduce students to real world research and ask participants to take on the role of apprentice researchers. Rather than being taught discreetly, language skills are developed in the context of a primary research project, and are “embedded in the nature of the activity rather than being taught in isolation or bolted on,” (Craig, 2011, P72). The research projects are team-based and once teams have identified and brainstormed (with appropriate instructor guidance) suitable topics for investigation, (recent topics have, for example, included issues of high school to university transition) they are given instruction in library search strategies and are asked to provide a short written overview of the selected source, complete with a live link for the instructor. Following class activities in summarizing, paraphrasing and quotation, the next task is to write a summary. Substantial feedback is provided before teams begin working on synthesizing their knowledge in a literature review. The process requires each team member to explain the content and relevance of their chosen article while team members take notes and ask for clarification. Once this phase of secondary research is completed, teams are asked to localize the global issue. A formal proposal is required in which students outline the aims and objectives of their intended primary research, as well as identifying and producing a suitable data gathering instrument. As the research develops, students are required, individually and in teams, to describe, explain and justify each stage, culminating in collaborative written reports, and oral presentations. Documentation is generated by the primary research process, with each new text building on and incorporating edited elements of previous documents and developing schema and knowledge from that process. Given that the focus is on developing students’ understanding, the need to turn to an external model to copy from is radically reduced and student ownership consequently increases. For example, one area where typically students tend to copy is the literature review. According to Levy and Ellis (2006), “Novice researchers tend to approach the literature review as nothing more than a collection of summaries of papers or an elaborated bibliography” (p182).

Our students, however, are required to select articles based on how the concepts, ideas and practices contained in them can inform their research. The review is therefore not a document in isolation, but rather integrated into the report as part of a *raison*

d'tre for decisions made in the research, such as informing methodological decisions and, as such, demonstrates the quality of students' critical thinking. Themes used to organize and construct surveys, for example, are typically derived from the review as are sub questions the research might focus on. The approach to the review (purpose) and the application of knowledge derived from it again reduces the need to plagiarize. It is therefore the nature of the tasks themselves which leads students to produce their own work. For instance, the recommendations section of the research report is approached first with a set of negotiated criteria as to what informs effective recommendations. These include identifying relevant stakeholders and beneficiaries, responsible parties who will apply the recommendations, appropriateness and applicability and so on, all of which is conditioned by the focus of the primary research of each team. It is simply not necessary (or really possible) to cut and paste a set of recommendations from some other context. Further, by this stage of the learning process, most students have a high level of buy in and ownership and are more concerned with originality.

This approach has by and large been, not only both in demonstrating critical thinking and understanding of content but also in significantly reducing plagiarism. This is largely due to the fact that each stage is carefully scaffolded which removes the need for copying, as task fulfilment cannot be achieved through reliance on "cut and paste" or lack of acknowledgement. There is no content as such to be memorized or copied, and any attempts to fool the reader are easily detected and challenged. Each member of the team is responsible for making sure there is no plagiarism, and in the process learns to describe each stage of their research, supporting their observations with primary data and reference to the literature. Students develop ownership, buy into the approach which simulates real world, adult, professional activity, learn that task fulfilment is rewarded and that not only is there no "need" to cheat, there is no opportunity.

Conclusion

The skills that are developed in these first year Communication courses can only be maintained if similar approaches are used throughout the curriculum. Expectations which are communicated to students through clear, achievable tasks and descriptions, transparent assessment and an inquiry based approach to learning would be a huge step in developing a deeper conceptual understanding required for higher order thinking. Other departments are beginning to follow suit in applying inquiry based approaches, and have adopted studio methods in the sciences. While no empirical data is available as yet, anecdotal evidence suggests that there is both less opportunity and need to cheat. Tasks are more concerned with documenting process and findings than with the memorization of content and students report being far more engaged than in the traditional methods often found in engineering schools.

While a clear interest in the issue is becoming evident in the region, future research into academic honesty at higher education institutions in the region could well focus on faculty perceptions and practice (particularly levels of tolerance) and their approaches to dealing with incidents. Other research questions might determine differences between departmental responses, between new and long-term faculty, and also differences between local and expatriate faculty.

References

- Akande, A. (1998) Towards the multicultural validation of a western model of students approaches to learning. *Education*, 119(1), 37–47. Available online: <https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=1267317&site=ehost-live&scope=site>
- Alvey, J.E. (2003) Adam Smith's Globalization (But Anti Secularization) Theory, Discussion Paper. Available online: <http://pur1.umn.edu/23716>. Accessed Jan 2012
- Beyer, PF. (1994) *Religion and Globalization* (Vol. 27) Sage Publications Limited, London
- Bjorklund, M. & Wenestam CG. (1999) Academic cheating: frequency methods and causes. Paper presented at the European Conference on Educational Research, Lahti, Finland., cited by Brennan, L. and Durovic, J. in *Plagiarism and the Confucian Heritage Culture (CHC) Student* Available: http://www.newcastle.edu.au/conference/apeic/papers_pdf/brennan_0512_edd.pdf
- Chickering, AW. and Gamson, ZF. (1987), Seven Principles for Good Practice in Undergraduate Education, *AAHE Bulletin*, 39 (1987), pp. 3–6. Available online: https://scholar.vt.edu/access/content/user/adevans/Public/DVDPPortfolio/Samples/samples/training/track_d/Introduction/Best%20Practices/Article%20-%207%20Principles%20of%20Good%20Practice%20in%20Undergrad%20Ed.pdf Accessed: March 2013
- Craig, R. (2011) Developing Cognition and Language Proficiency through the Acquisition and Articulation of Knowledge: Real World Communication Activities for Engineering Students in and across the Disciplines, *International Journal of Arts and Sciences*, Vol 4, no. 10, pp 69–76
- Daily Telegraph, (19th May 2012). *More students caught copying university applications*. Online at: www.telegraph.co.uk/education, Accessed June 2012
- Faculty and Academic Integrity (1997) *Synthesis: Law and Policy in Higher Education*, Gary Pavela, editor. Available: <http://www.academicintegrityseminar.com/Teaching/TenPrinciples.html>
- Gulf News (2012) *Student cheating is a serious problem in the UAE, academic says*. Available online: <http://gulfnews.com/news/gulf/uae/education/student-cheating-is-a-serious-problem-in-the-uae-academic-says-1.1061287>, Accessed Feb 2012
- Hatherley-Greene, P. The reasons Emirati men struggle at university, *Special to Gulf News*. September 1, 2012c. Available online at: <http://gulfnews.com/news/gulf/uae/education/the-reasons-emirati-men-struggle-at-university-1.1066944>. Accessed: June 2012
- Hofstede, GH. (2001) *Culture's consequences* (2nd ed.). Thousand Oaks, CA: Sage, London
- Ivy League cheaters: Harvard University investigating 125 undergrads “for plagiarizing test answers” *Daily Mail Online* (Aug 2012) at: <http://www.dailymail.co.uk/news/article-2195987/Half-Harvard-class-cheated-exam-University-investigating-125-undergrads-plagiarizing-answers.html>. Accessed June 2012
- Khaled, A. (2008) Egypt: Research plagued by plagiarism. *University World News*, Issue No 9: Issue 00246, 20 July 2008. Available online at: <http://www.universityworldnews.com/article.php?story=20080717165104870&query=Research+plagued+by+plagiarism>, Accessed Feb 2012
- Levy, Y. & Ellis, JT. (2006) A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science Journal* 9, 2006 pp 181–202. Available online: <http://www.scs.ryerson.ca/aferworn/courses/CP8101/CLASSES/ConductingLiteratureReview.pdf>. Accessed March 2012
- McCabe, DL. (2005) Cheating among college and university students: A North American perspective. *International Journal for Educational Integrity*, Vol. 1, no. 1, 10–11. Available online: <http://www.ojs.unisa.edu.au/index.php/IJEI/article/view/14/9> Accessed: March 2012

- McCabe, DL. & Pavela, G. (2004) Ten (Updated) Principles of Academic Integrity. How Faculty Can Foster Student Honesty. *Change*, Vol. 36, No. 3, pp. 10–15. Available online: <http://www.jstor.org/stable/40177967>, Accessed: March 2012
- Petroleum Institute, Mission Statement [Online]. Available from: http://www.pi.ac.ae/PI_INS/eo/info/mission.php. Accessed 16 January 2013
- Report of the Josephson Institute of Ethics (2009) A Study of Values and Behavior Concerning Integrity: The Impact of Age, Cynicism and High School Character. Available: <http://josephsoninstitute.org/surveys/> Accessed Apr 2012
- Rettinger, DA. (2005) The Relations among Religion, Motivation, and College Cheating: A Natural Experiment. *Ethics & Behavior*, 15(2), pp 107–129. Lawrence Erlbaum Associates
- Russouw, R. (2005) Net closes on university cheats. *Saturday Star*. Retrieved October 9, 2006, from http://www.iol.co.za/index.php?set_id=1&click_id=13&art_id=vn20050226112531241c573402. In Batane, T. (2010). Turning to Turnitin to Fight Plagiarism among University Students. *Educational Technology & Society*, 13 (2) pp 1–12
- Sheard, J., et al. (2002) Cheating and Plagiarism: Perceptions and Practice of 1st Year IT Students. *Proceedings of the 7th Annual Conference on Innovation and Technology in Computer Science Education*. Pages 183–187. ACM SIGSE bulletin (Vol. 34, no. 4)

Authors

Robert Craig, rcraig@pi.ac.ae,
David Dalton, ddalton@pi.ac.ae
Petroleum Institute, Abu Dhabi, United Arab Emirates

Copyright © 2013 Authors listed on page 103: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

TEACHING STAFF CONCERNS ABOUT ACADEMIC INTEGRITY AND THEIR IMPLICATIONS FOR STAFF DEVELOPMENT

Sharon Flynn

Abstract: Anecdotal evidence suggests that, even when academic staff are aware of the existence of a plagiarism policy, many still don't use it. They either prefer to deal with it in their own way, or ignore the issue. In order to encourage increased and improved use of the existing policy, we were interested in understanding staff concerns about plagiarism generally. As part of a Postgraduate Certificate (PG Cert) in Teaching and Learning in Higher Education, a 3 hour workshop is devoted to Academic Integrity. At the start of the workshop, participants are asked to complete a survey on their experience and concerns regarding plagiarism, and their awareness of local policies and procedures. Staff are then asked to articulate their single biggest concern around academic integrity, and this collection of concerns and the ensuing discussion is used to direct aspects of the workshop. With different groups, the emerging group concerns will have different flavours; some teaching-focused, some student-focussed and some policy-focussed. The data from the surveys over the last 3 years gives an insight into staff awareness of policies and procedures at our institution as well as highlighting the general ethos and attitudes regarding student plagiarism. This paper will compare the collected data against early results from the IPPHEAE project staff survey. Knowledge about the levels of awareness of plagiarism policy amongst staff has implications for staff development. This paper will give an overview of various approaches to awareness raising and staff development that have been used in our institution, with an indication of the levels of success and factors contributing to this.

Introduction

The National University of Ireland, Galway, is a university in the west of Ireland, with about 17,000 students and more than 2,000 staff across 5 Colleges, including Arts & Humanities, Business & Law, Science, Engineering, and Medicine & Nursing.

The Centre for Excellence in Learning and Teaching (CELT) is a central unit of the university which provides support to academic staff for teaching and learning as well as playing a role in determining policy in the areas of learning, teaching and assessment. One aspect of CELT's remit is to provide a Postgraduate Certificate (PG Cert) in Teaching and Learning in Higher Education, offered to academic staff. Each year, about 36 staff enrol on this 30 ECTS programme, which is voluntary. In any year, the participant profile ranges across all disciplines, with a range of years of experience. Any staff member who completes the PG Cert can progress to a Postgraduate Diploma (PG Dip) in Academic Practice and, ultimately, an MA in Academic Practice.

The university takes a holistic approach to plagiarism, informed by good practice in the UK, the United States and Australia. In 2004, a Code of Practice for Dealing with Plagiarism was developed, which established the role of *plagiarism advisor*—a member of academic staff, associated with a School, whose role is to deal with reported cases of plagiarism within that School according to the code. In 2006, the *plagiarism committee* was created, made up of plagiarism advisors, and giving them the authority to apply

penalties according to an agreed set of criteria. In 2012, the code of practice (National University of Ireland, Galway, 2012) was reviewed and revised, informed by experience and incorporating a version of the Benchmark Plagiarism Tariff produced by the third phase of the Academic Misconduct Benchmarking Research (AMBeR) project (Tenant and Rowell, 2010).

Since 2006, the Turnitin originality detection tool has been used, in varying ways, across the university and in 2008 it was integrated into the university's virtual learning environment (VLE), allowing it to be used with large numbers of students with a minimum of administrative overhead (Flynn, 2010). Associated with the technology, regular workshops on using Turnitin to deter plagiarism are offered to all academic staff. Support on using the software using a proactive and constructive approach is also offered.

General advice on aspects of academic integrity, interpretation of the Code of Practice, and approaches to deterring plagiarism is available on demand. However, there are still concerns about the level of staff awareness of the institutional policy and its implementation, and awareness of approaches to addressing issues of student plagiarism generally.

A Workshop on Academic Integrity

To address some of these concerns, in 2009 it was proposed to incorporate a workshop on academic integrity into the PG Cert module on Course Design, Assessment and Evaluation. The module is given as a series of seven 3 hour workshops on aspects such as curriculum design, learning outcomes and Bologna, student engagement, evaluating teaching, and assessment. Critically, the workshop on Academic Integrity is scheduled to follow the workshop on Assessment, so that participants can make links between good assessment practice and plagiarism deterrence.

The format of the workshop is described in the following paragraphs.

Taking time to focus

Participants (approx 16 in each group) are asked to complete a survey, asking about their knowledge of the code of practice; their level of concern about plagiarism in the classes they teach; the number of cases they have seen in the last year; the name of a plagiarism advisor for their School; who they would ask for advice about student plagiarism; whether they discuss issues of plagiarism with students, tutors or other staff; and finally, what tips they have for deterring plagiarism. The purpose of the survey is to focus them on their own thoughts around issues of academic integrity, prior to the workshop discussion. Participants are informed that the survey data may be used as part of research.

They are then asked to write down their single biggest concern about plagiarism. After some minutes, participants are asked to discuss with their neighbour what they have written down. This usually creates a good buzz in the room and people begin to relax a little. Finally, the facilitator goes around the room, asking each person to describe their concerns, which are noted on a flipchart or whiteboard. Links can be established and items clustered to get an overall picture of those issues of most concern

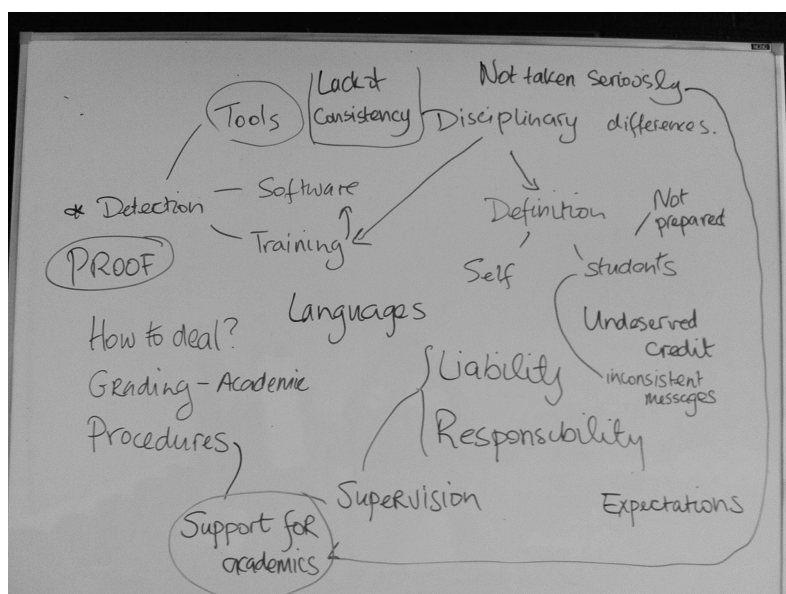


Figure 1. Staff concerns, group A, 2012

to the group. See, for example, figure 1, showing the result of this exercise for one group in March 2012.

This discussion of staff concerns serves a number of purposes. First, it allows each participant to articulate some aspect of academic integrity that is important to them. Second, it allows the facilitator to focus the workshop on the collected issues, particularly where there may be misconceptions. By often referring back to the collection of concerns during the three hour workshop, participants feel a sense of ownership of the discussion. Finally, it gives the facilitator a very useful insight into staff perceptions of academic integrity which, over three years, have built into a valuable collection of data.

The workshop continues

The workshop continues with a discussion of what we understand by the term plagiarism. In particular, we use the *Where do you draw the line?* exercise from Carroll (2007). This usually demonstrates that there is inconsistency in staff views of plagiarism, even when staff come from the same discipline and teach the same groups of students.

From this realisation, we move to consider the student view of plagiarism, why students plagiarise, and what we might do as academics to address these issues.

The middle hour of the workshop is spent looking at internet plagiarism, including a tour of ghost-writing sites, essay mills and other web-based opportunities available to students. This is often very shocking for the group, since they may have never seen or considered the availability of such services for students.



Figure 2. Wordle representing Staff concerns, 2010

The group then focuses back on good practice for assessment and how this can be used to deter plagiarism. This links nicely with the previous workshop on assessment, and returns the general mood to one of positivity and empowerment.

Finally, we consider policy and what makes a good policy. We discuss reasons why academic staff might choose not to follow the policy, including administrative overhead or lack of trust.

Assessment

The overall assessment of the module is based on the development of a course review folder, focusing on one course that is currently being taught, and reviewing it in the context of the seven workshops.

For the workshop on academic integrity, participants are asked to consider an existing student assignment, analyse it and identify opportunities for plagiarism, make changes to design-out opportunities for plagiarism. They are also asked to consider how they might use the opportunity to help students learn about issues of academic integrity.

Academic Staff Concerns

The workshop, as described in the previous section, has now been running annually as part of the PG Cert in Teaching and Learning in Higher Education for 3 years, with another workshop due to run in March 2013. Staff feedback has been very positive.

Data about academic staff concerns, collected as described previously, has been retained as flip chart pages or photographs taken of the final whiteboard. The records are also made available to the participants, via the VLE, after the workshop, sometimes using a graphical form (see figure 2).

The data over three years has been analysed using a simple keyword approach and grouped according to whether the concerns tend to be centred on students, matters concerning teachers and teaching styles, issues of policy, or concerns related to the institution.

In this section, we describe those issues of primary concern to academic staff under four headings. In the next section, we will consider how awareness of policies and procedures, according to our survey data, compares to the early results of the staff survey for the project Impact of Policies for Plagiarism in Higher Education Across Europe (IPPHEAE) (IPPHEAE, 2013). Finally, we will discuss the implications for staff development at the National University of Ireland, Galway.

Student focused

Student focused concerns can be grouped into five categories:

- Students' lack of skills and awareness (S.1)
- Opportunity for plagiarism (S.2)
- Effect on learning (S.3)
- Impact for training (S.4)
- Reputation (S.5)

One of the key concerns expressed by staff is that students lack the skills necessary to avoid plagiarism (S.1). There is a strong belief that students are plagiarizing innocently or unintentionally, that they do not have a good sense of what is expected of them in higher education. Connected to this is the observation that students do not have the skills necessary to complete written assessments in higher education, including information literacy skills, study skills as well as basic writing skills. Coming from secondary school, in particular, the concern is that students are not well prepared for learning. When information about plagiarism and plagiarism avoidance is made available, through induction or course guidelines for example, often students become confused or worried about their own skills, which may further compound the problem.

Related to this is that students often receive mixed messages about what is expected of them and what is acceptable when it comes to academic writing. One teacher might be very relaxed towards acknowledging the work of others, while another teacher is very strict about particular referencing conventions. Many students do not realize how serious plagiarism is.

The second area of concern to staff is the increased opportunity for students to plagiarise, sometimes related to the opportunity offered by technology (S.2). In general, staff refer to the tendency for students to use a copy and paste approach to writing; they are not (yet) aware of essay mills and ghost writing services. This concern is linked to the first (S.1) in that students have developed this, previously successful, approach to writing and believe it to be an appropriate strategy at university level.

Staff who are involved in teaching large cohorts of students, in some cases up to 850 students in one class group, will identify collusion, or inappropriate collaboration, as an issue. Collusion often occurs between students in the same class group, or between years in the same overall programme.

A small number of staff will identify laziness as an issue: some students don't want to put the effort into an assessment and will resort to other approaches.

One particular issue that comes up occasionally is that of language/translated plagiarism, most often from language teachers. This can centre around technology and the use of tools such as Google Translate, for example. There are also anecdotes about students using their friends, who may be native speakers of the language to be learned, to complete assignments for them.

A significant number of staff suggest that plagiarism is an impediment to learning (S.3), and that by plagiarizing the student is missing an opportunity to learn and to develop her own ideas. Connected to this is the possible impact of widespread plagiarism on hard-working students, and that their learning may be affected by the actions of others.

The lack of awareness and skills has an impact on teaching (S.4). Academic staff are concerned that there is a need for training for students about what is plagiarism and how they can avoid it, information skills and referencing/citation skills. It is not clear, from the ensuing discussions, who is expected to provide such training.

And finally, staff are worried that the student who plagiarises is given undeserved credit if the act goes unnoticed. On the other hand, if the case is reported, they worry about the reputation of the student (S.5). Where is the record kept and who has access to it? This is particularly relevant if the student is taking a course leading to a professional qualification, for example Nursing.

Teacher focused

Often, the concerns of the group, as articulated, focus on the teacher and how academic dishonesty can have an effect on the role of the teacher. These are grouped into four categories:

- Time and effort (T.1)
- Impact on teaching and assessment (T.2)
- How to deal with issues (T.3)
- What support is available (T.4)

One of the primary concerns raised by workshop participants is the time and effort required to actively monitor and address plagiarism (T.1). In an environment where there is increasing pressure to produce more research and teach more students, monitoring of student activities in large groups is not insignificant. Where problems with writing are found, resulting from either intentional or unintentional plagiarism, the time and effort required to follow up, collect evidence, and go through the formal process, can be off-putting. For many staff members, particularly those who have had a bad experience in the past, it is simply easier to ignore issues of plagiarism and just reduce the mark of the student. This, in turn, has an effect on the culture of the organisation (U.1).

Another concern is that the issue of plagiarism restricts what can be done with continuous assessment (T.2). Staff worry that they will have to resort to final, written exams. While staff generally agree that continuous assessment supports student

learning, this only works if the expectations of the staff and students match, and if students genuinely put the work into the assessment.

A second issue around the effect on assessment is that of grading. Linked to (S.5), students may be getting undeserved credit for unoriginal work, so the opportunity for learning has been missed. Alternatively, staff might be dealing with the issue by simply marking students down, where an opportunity to learn about writing and referencing skills may be missed.

Some participants will be concerned about what to do if/when plagiarism is discovered (T.3). Some are not aware of the code of practice, while others may have had discouraging experiences while using it in the past.

Finally, there are many questions around what support is available to staff, either within a discipline or at institutional level (T.4). Some staff will have a strong sense of support from their colleagues, while others may feel unsupported within their School or discipline, perhaps referring to an unsympathetic head of unit.

Policy focused

Policy focused concerns are grouped into three categories

- Penalties (P.1)
- Fairness (P.2)
- Usage (P.3)

Concerns around penalties (P.1) tend to focus on whether they are appropriate, with some staff believing that existing penalties are too strong and others believing that they are too weak. A major concern is whether they are consistently applied across the university. Clarity is needed around the criteria used to determine penalties and how borderline decisions are made.

Fairness is raised as an issue in a number of contexts (P.2). Is the policy inherently fair? Is the policy being applied in a fair and consistent manner? Is it fair to all students and staff? For example, if the penalty applied for a (first) plagiarism offence is simply a warning, is that fair to other students who have not plagiarized?

There is also a concern that the policy is not being used by all staff (P.3), through ignorance of its existence, a choice based on time and effort required (T.1), or a lack of trust in the process (U.1). This, in turn, has implications for the fairness of the policy (P.2).

Institution focused

Finally, the institution focused concerns are grouped into two categories

- Culture (U.1)
- Reputation (U.2)

Concerns about the culture of the institution or the discipline (U.1) come up in different ways. Some participants observe that there is a culture of ignoring the issue of academic dishonesty among students, or not taking it seriously, perhaps in the hope

that it might go away. Others observe a general lack of trust in the policy or a general reluctance to use it. In some disciplines there is a reluctance to even discuss the issues, and individual staff members can feel very isolated through lack of support (T.4).

The concern about institutional reputation (U.2) is clear. The integrity of the qualification awarded as well as the integrity of the institution is under threat and most staff have some worries about the accreditation of students who have not met requirements.

Awareness of Policies and Procedures

As part of the initial survey undertaken by participants at the workshop, they are asked about the level of awareness of the Code of Practice for Dealing with Plagiarism. Specifically, the following questions are included:

- Are you aware of the NUI Galway Code of Practice for Dealing with Plagiarism?
- Have you read the NUI Galway Code of Practice for Dealing with Plagiarism?
- Can you name a plagiarism advisor for your School? Enter name in the text box.

In the first year of the survey, 2010, almost 87% of respondents were aware of the existence of the policy, with a slight variation across different Colleges. However, only 60% of respondents had read the document, and just over half (52%) could correctly name a plagiarism advisor for their school. Given that the group was made up of academic staff who, by volunteering to take the PG Cert, could be said to have an increased awareness and interest in issues around teaching and learning, we assume that these figures are higher than across the general University community.

By March 2012, the percentages had increased, with all survey respondents reporting an awareness of the policy and over 80% admitting to have read it. In the intervening time, the role of plagiarism advisor had been strengthened and made more visible. Only one person (from 29) was unable to name a plagiarism advisor for his School.

This tallies with very early data from the IPPHEAE project, relating specifically to NUI Galway. In late 2012, teaching staff at NUI Galway were made aware, via email, of the IPPHEAE project and encouraged to take the staff survey available at the project information website ippheae.eu.

From the raw data provided by IPPHEAE for NUI Galway (IPPHEAE, 2013), 12 teaching staff members at NUI Galway responded to the survey. In response to the questions:

- 5.b. This institution has policies and procedures for dealing with plagiarism
- 5.c I believe this institution takes a serious approach to plagiarism prevention
- 5.d I believe this institution takes a serious approach to plagiarism detection

all twelve responses are either “Agree” or “Strongly Agree”.

This is encouraging and indicates a good level of awareness of the existence of the code of practice and its use.

Less encouraging, perhaps, are the responses to the questions:

- 5.g. Penalties for plagiarism are administered according to a standard formula

Table 1

Results from IPPHEAE Staff Survey

Question	Strongly Agree	Agree	Not sure	Disagree	Strongly Disagree
5.g	6	3	3		
5.h	6	1	3	1	1
5.i	6	2	4		

- 5.h. I know what penalties are applied to students for different forms of plagiarism and academic dishonesty
- 5.i. Student circumstances are taken into account when deciding penalties for plagiarism

which demonstrate that while staff may be aware of the policy, they are not so familiar with the content or associated procedures (see Table 1).

Further evidence that teaching staff are not familiar with the content of the NUI Galway code of practice, and in particular the role of the School plagiarism advisor, can be found in the answers provided to the questions:

- 15.a. Who decides whether a student is guilty of plagiarism?
- 15.b. Who decides on the penalty applied to students for plagiarism?

In the data collected, only two respondents (17%) mention the plagiarism advisor as the person responsible for making these decisions. Other responses vary from the individual tutor, a department panel or an institutional panel.

The number of responses to the IPPHEAE is very small, just 12 out of a possible 2000, and therefore not statistically valid. However, the responses come from staff across 4 of the 5 Colleges and they do support the results of the workshop surveys. The conclusion is that while academic staff are aware of the existence of university policies and procedures to deal with plagiarism, they are not familiar with the content.

Consequences for Staff Development

It is clear, from the survey data and from experiences in supporting academic staff, that staff development in the area of Academic Integrity is needed. Academic staff have a range of concerns, many of which can be addressed through training opportunities, provision of resources and the availability of advice and support.

Approaches that have been taken at NUI Galway to provide training opportunities, along with an indication of success, are described in the following paragraphs.

Workshops as part of PG Cert

This workshop is as described in previous sections. The advantage of the workshop is that it is embedded firmly within the context of teaching and learning, and relates explicitly to good practice in student assessment. The concerns of staff are addressed directly through the format of the workshop and any questions or misconceptions about policy and procedure are addressed.

Feedback from the workshop has been encouraging. Sample statements from participants include:

- Until the class devoted to the theme of academic integrity/ plagiarism, it was a case of don't ask, don't tell. (Participant A)
- Ultimately, the best way to deter plagiarism is to develop and encourage student centered (sic) learning that encourages academic integrity. (Participant B)
- It raises so many issues and concerns in my head, not just for oneself or other staff members designing assignment and unknowingly or knowingly receiving plagiarized work, but also for the students. (Participant C)

Participants come from a range of disciplines, so it can be expected that these individuals may go back to their respective units and start to address the culture from within. However, it is a small number of individuals, about 80 to date, and a complete change of culture will take a long time.

Technology Workshops

These are workshops based around the use of a particular technology, for example Turnitin or GradeMark, and offered to a departmental group or to any interested staff. While the focus is on the technology, these offer an opportunity to discuss general issues of academic integrity.

Our experience has been that these workshops are most successful when offered to a departmental group, since discussions can be encouraged concerning a consistent approach to dealing with academic integrity within the discipline. As noted in Flynn (2010), this encourages a shared understanding of academic integrity across a teaching team, resulting in a consistent message for students.

Workshops on Deterring Plagiarism

A version of the PG Cert workshop, entitled Deterring Student Plagiarism is offered on request to discipline based groups. Like the technology workshops, this works best when staff from a single discipline or teaching team come together and can discuss issues in a constructive and supported environment. Again, this leads to increased consistency and clarity for students.

A workshop of this type has the potential to transform the culture within a discipline, but requires commitment from the participating staff and strong leadership to drive the change.

Availability of Resources and the Role of the Plagiarism Advisor

It will not be a surprise to anybody working in academic development that teaching staff find it difficult to find time to come to training workshops. Academic staff usually attend a workshop because it forms part of a wider programme (as in the PG Cert), the staff member has a particular need (as in the technology workshop) or the staff member has been compelled to attend by a head of unit.

Where resources are available, teaching staff will not discover these until such time as they are needed, usually because of a problem with plagiarism within the teaching context.

We suggest that, as the role of the plagiarism advisor becomes more embedded within the School or discipline organisation, there is potential to further promote values of academic integrity across all university activities, including teaching, learning, supervision and research.

Related Research

The literature recognises the need for a holistic approach to plagiarism prevention, where there is a shared responsibility between students, staff and institution (Macdonald and Carroll, 2006).

Carroll (2007) recommends that staff should be informed about the policy for plagiarism, to ensure everyone knows about it, and new staff should be informed as part of their induction. Anecdotal evidence suggests that, even where staff are aware of the existence of a policy, many still don't use it. They either prefer to deal with it in their own way, or ignore the issue.

Flint (2004) explored the staff perspective on plagiarism and showed that there was a strong emotional aspect to dealing with plagiarism, resulting in a variety of strategies being used. The paper concludes that an understanding of the issues could be used to inform future actions, including staff development.

Other authors (for example Bennett et al, 2011) have considered faculty perceptions of and attitudes towards plagiarism, demonstrating a lack of consistency in definitions and differences in perceptions between staff and students. Pickard (2006) suggests that better understanding of staff and student perceptions towards plagiarism will be useful in guiding staff development. Andrews et al (2007) identify three themes in narrative comments from staff surveys, relating to factors that contribute to cheating behaviour, lack of support for staff in reporting plagiarism, and institutional policies. Zivcakova et al (2012) in a study involving 8 faculty members, found that faculty concerns could be categorised into two sub-themes: issues with other faculty members, and institutional policy issues.

Conclusion

In this paper we demonstrate that teaching staff have a wide range of concerns regarding academic integrity and how it affects students, staff and the institution as a whole. Some of these concerns can be addressed through staff development and the provision of support for staff and students.

While the Academic Integrity workshop as part of the PG Cert produces advocates with an awareness of the issues, they are still individuals within larger units.

Key to the success of staff development efforts is a willingness to address the culture of the unit, discipline or institution, and a realisation that values of academic integrity have to be embedded across all teaching and learning activities.

References

- Andrews, K., Smith, L., Henzi, D. & Demps, E. (2007) Faculty and Student Perceptions of Academic Integrity at U.S. and Canadian Dental Schools', *Journal of Dental Education*, 71(8), 1027–1039
- Bennett, K., Behrendt, L. & Boothby, J. (2011) "Instructor Perceptions of Plagiarism: Are We Finding Common Ground?", *Teaching of Philosophy*, 38(1), 29–35
- Carroll, J. (2007) *A Handbook for Detering Plagiarism in Higher Education*, 2nd edition. Oxford Centre for Staff and Learning Development
- Flint, A. (2004) "Emotion, Practice, and Plagiarism: Exploring the Staff Perspective". *Plagiarism: Prevention, Practice and Policies*, 28–30 June 2004, Newcastle upon Tyne, UK
- Flynn, S. (2010) "Using Turnitin with large classes to support writing". *Fourth International Plagiarism Conference*, 21–23 June 2010, Newcastle upon Tyne, UK
- IPPHEAE (2013) Impact of Policies for Plagiarism in Higher Education Across Europe, European Union Lifelong Learning Programme (510321-LLP-1-2010-1-UK-ERASMUS-EMHE (2010-3491))
- Macdonald, R. & Carroll, J. (2006) Plagiarism—a complex issue requiring a holistic institutional approach, *Assessment & Evaluation in Higher Education*, 31:2, 233–245
- National University of Ireland, Galway. (2012) Code of Practice for Dealing with Plagiarism [Online]. Available from: www.nuigalway.ie/plagiarism [Accessed 29 January 2013]
- Tennant, P. and Rowell, G. (2010) Benchmark plagiarism tariff [Online]. Available from: www.plagiarismadvice.org/images/plagad/resources/Tennant_benchmarkreport.pdf [Accessed: 29 January 2013]
- Zivcakova, L., Wood, E., Baetz, M. & De Pasquale, D. How do faculty members respond to their students' discussions of academic misconduct and academic integrity? *International Journal for Educational Integrity*, 8(1), 25–40

Acknowledgements

Part of the data used in this study was collected for the project Impact of Policies for Plagiarism in Higher Education Across Europe (IPPHEAE), led by Irene Glendinning of Coventry University, UK, with funding from the European Union Lifelong Learning Programme (510321-LLP-1-2010-1-UK-ERASMUS-EMHE (2010-3491)).

Author

Sharon Flynn, sharon.flynn@nuigalway.ie, National University of Ireland, Galway, Ireland

Copyright © 2013 Sharon Flynn: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

DO STUDENTS THINK WHAT TEACHERS THINK ABOUT PLAGIARISM?

Tomáš Foltýnek, Jiří Rybička, Catherine Demoliou

Abstract: Preventing students' plagiarism is an important issue at Higher Education Institutions (HEI). Teachers who are primarily involved in plagiarism prevention have to explain to students what plagiarism is and to teach them how to avoid plagiarism. A survey conducted under the research project on the "Impact of policies for plagiarism in higher education across Europe" (IPPHEAE) has collected thousands of questionnaires from both students and teachers of several HEI at various EU countries. The project has addressed several research questions relating to plagiarism. This paper compares and contrasts the overall data collected on students' and teachers' attitude to academic writing and their perception of plagiarism and plagiarism penalties as well as their knowledge of Institutional policy/procedures on plagiarism. On the basis of the results obtained it would appear that changes in teachers' attitude and a better understanding of students' educational needs and perception of plagiarism may provide the means to enable HEI to help students to avoid plagiarism in a more effective way.

Introduction

Plagiarism is becoming an important issue of increasing concern in our societies these days (Borg, 2009). Views and attitudes on plagiarism, however, differ from culture to culture as shown by the study of Carroll (2008) on foreign students going to UK, and by the study of Sutherland-Smith (2008) on international students going to Australia. Plagiarism is an interdisciplinary issue that combines psychological and sociological aspects with ethical and legal ones, and has a national and cultural context (Sutherland-Smith, 2005). Although most of the western European countries are sensitive to plagiarism there are countries, like those in Eastern Europe, where plagiarism is not considered to be a big problem (Foltýnek and Čech, 2012). As a consequence plagiarism is most often judged on the basis of what is considered correct in the "western" societies without taking into account any country specific cultural roots (Haynes and Introna, 2005).

Several studies have already been published on students' and teachers' perception of plagiarism. Risquez et al. (2013) found that although students declare a personal interest in plagiarism and consider plagiarising to be generally bad and punishable, they fail to distinguish specific cases of plagiarism. In fact, in most cases, students are unable to judge whether there is a case of plagiarism or not. In addition, neither students nor teachers appear to be consistent when expressing themselves as to what they consider as plagiarism. In fact, as Gu and Brooks (2008) have shown, students may actually express their teachers' views rather than their own, thus making it difficult to pinpoint any differences in students' views. Furthermore, many students may omit citing and referencing because they are not able to recognize where it is required to do so (Risquez et al., 2013), thus committing plagiarism unintentionally. Regarding teachers, they usually do not punish plagiarism intentionally. Most often they take

plagiarism as their fail, and either they do not look for cases of plagiarism or they choose to ignore them (Sutherland-Smith, 2005).

The majority of students are thought to become aware of plagiarism before or when they start their bachelor degree studies. Although some students may know how to cite and to reference literature sources correctly when they start university, these skills are really taught and practiced during their undergraduate studies (Foltýnek and Čech, 2012). University faculty are expected to provide the opportunity to students to develop their skills of writing assignments and research projects/theses without plagiarising. However, this may not be possible if there are differences between teachers' and students' perception/awareness of (a) what constitutes academic writing and the reasons that students resort to plagiarism, (b) the type of penalties imposed in cases of plagiarism, and (c) the institutional policies and procedures that deal with plagiarism. This study was aimed to identify if teachers and students from HEIs of several EU-countries differ significantly in their perception and awareness of the above. Understanding such difference may enable the development of a European wide plagiarism prevention approach through education, with teachers as the protagonists.

Material and Methods

We have examined the answers to a survey conducted under the IPPHEAE project with a focus on plagiarism policies, procedures, prevention and penalties at various HEI in EU countries across Europe. About 2588 and 572 questionnaires from students and teachers, respectively, were collected namely from UK, Poland, Lithuania, Czech Republic, Cyprus, Slovakia, Germany, Greece, Bulgaria, Portugal, Austria, France, Finland, Malta and Ireland.

For the purpose of this paper, only some of the questions and statements were analysed. Specifically, we have compared the following:

- I have received / Students receive training in techniques for scholarly academic writing and anti-plagiarism issues
- This institution has policies and procedures for dealing with plagiarism
- Plagiarism policies, procedures and penalties are available to students
- Penalties for plagiarism are administered according to a standard formula
- Student circumstances are taken into account when deciding penalties for plagiarism
- I think that translation across languages is used by some students to avoid detection of plagiarism
- What would happen if a student at your institution was found guilty of plagiarism in their assignment or final project / dissertation?
- Is there any referencing style you / students are required or encouraged to use in written work for your course?
- Which of the following services are provided at your institution to advise students about plagiarism prevention?
- What do you / students find difficult about academic writing?

- What leads students to decide to plagiarise?
- Assuming that 40% of a student's submission is from other sources and is copied into the student's work as described in (a–f) below, indicate your judgment on plagiarism
 - a) word for word with no quotations
 - b) word for word with no quotations, has correct references but no in text citations
 - c) word for word with no quotations, but has correct references and in text citations
 - d) with some words changed with no quotations, references or in text citations
 - e) with some words changed with no quotations, has correct references but no in text citations
 - f) with some words changed with no quotations, but has correct references and in text citations

For the purpose of clarity, the above were divided into three areas of concern within which the students' and teachers' responses were analysed accordingly.

Results and Discussion

Training in techniques for scholarly academic writing

The majority of EU students and teachers agreed (62%) that students receive training in techniques for scholarly academic writing and anti-plagiarism issues during their studies. There were differences between EU countries in the percentages of positive answers given. However, since the answers between teachers and students from the same HEI were in agreement, it became apparent that there are institutions in EU countries, which do not provide students with any training on plagiarism.

The analysis of data on how students learn/get trained on academic writing that will enable them to avoid plagiarism (Fig.1), showed a student preference (54%) for the web. This source of learning was underestimated by teachers who chose instead their own methods (i.e. class lectures and guidance notes), as the most preferred by students to help them avoid plagiarising.

Similarly, when participants were asked to choose what students may find most difficult about academic writing, the majority of students chose different aspects to those considered by teachers as most difficult (Fig. 2). For example, whereas the majority of students (64%) indicated that finding a good quality source for reading material was the most difficult aspect of academic writing, the majority of teachers (92%) thought that students have more difficulty in understanding the different referencing formats.

Both of the above paradigms indicate that teachers need to pay greater attention to students to identify their needs and expectations when it comes to writing academic work. Based on the above, one may then assume that if students received a better training in academic writing they should be able to write assignments without having to resort to plagiarism. However, would training in academic writing alone be enough to prevent plagiarism?

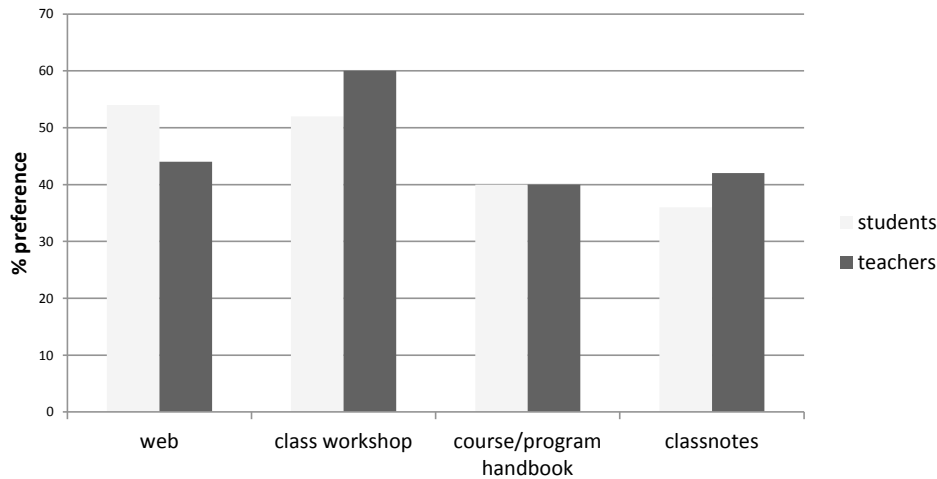


Figure 1. Differences in the preference for teaching and learning sources for educating students in academic writing

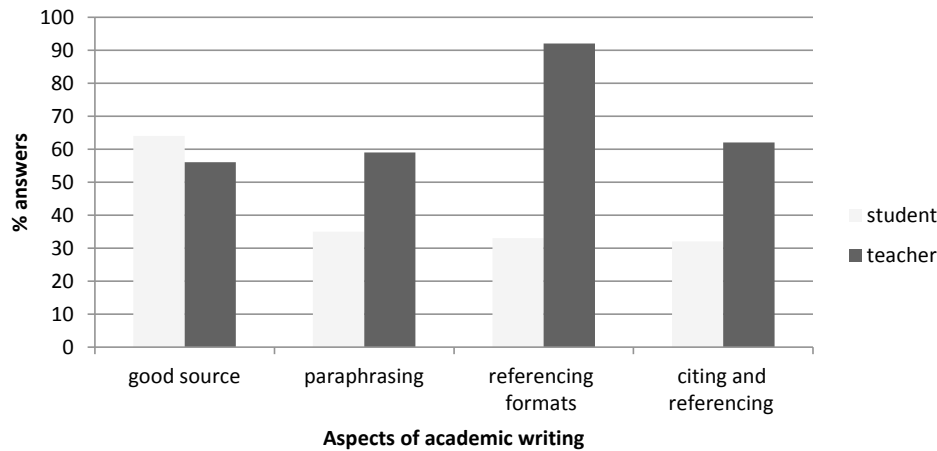


Figure 2. Differences in the perception of what students and teachers consider as difficult in academic writing

Plagiarism prevention

In trying to prevent plagiarism effectively, one also needs to understand the reasons students plagiarize. Both teachers and students were given the same set of 21 reasons for students plagiarizing, and asked to choose the 10 most likely ones. When the data was analysed in terms of choice differences between teachers and students, it became apparent that teachers underscored a number of reasons that students consider as the

Table 1

Most likely "reason for students plagiarizing"

Reasons	Chosen by teachers (%)	Chosen by students (%)
It is easy to cut and paste from the Internet	83	67
Plagiarism is not seen as wrong	53	32
They think the lecturer will not care	46	32
Their reading comprehension skills are weak	39	26
They don't see the difference between group work and collusion	28	16
They run out of time	52	64
They are unable to cope with the workload	24	41
They think their written work is not good enough	20	30
They feel the task is completely beyond their ability	15	28
Assignments tasks are too difficult or not understood	10	23

cause for resorting to plagiarism. Furthermore, teachers overscored the importance of other causes.

As shown in Table 1, both teachers and students showed similar trends in the order of what they considered as important reasons for plagiarizing in agreement with those reported by DeVoss and Rosati (2002). According to the teachers' opinions, students plagiarize mostly because they are lazy, lack skills of reading comprehension, they think nobody will care and are unable to recognize that plagiarism is wrong (Table 1). However, according to students' choices, it appears that running out of time, inability to cope with the workload and deadlines and a sense of insecurity are the most likely reasons for students to plagiarize. The apparent problem with time management as indicated from these results may explain why students choose to resort in cut and paste from the Internet when writing an assignment as indicated by the most popular reason for plagiarizing that was chosen by students (Table 1). The underscoring of the above reasons by teachers, would suggest that teachers may differ in their perception of the difficulty of a task/assignment given to students. Looking at the two reasons, i.e., "the task is beyond their ability" and "tasks are too difficult or not understood" (Table 1), chosen by about 25% of the students but only by 10–15% of teachers, one could conclude that if assignments/tasks were explained better, students may be able to complete these to a much higher standard without resorting to plagiarism. However, as McDowell and Brown (1998) suggest, this may not be the case since teachers and students judge the difficulty of assignment tasks differently. The present results suggest that a better approach in preventing students from resorting to plagiarism may be by getting them to manage their time better and by providing them with sources for background reading to help them understand the assigned task. Such an approach may also motivate students more and make them less afraid to deal with it.

However, students may still plagiarize if there is no consensus between teachers and students on what constitutes plagiarizing. In order to check how close are the teachers'

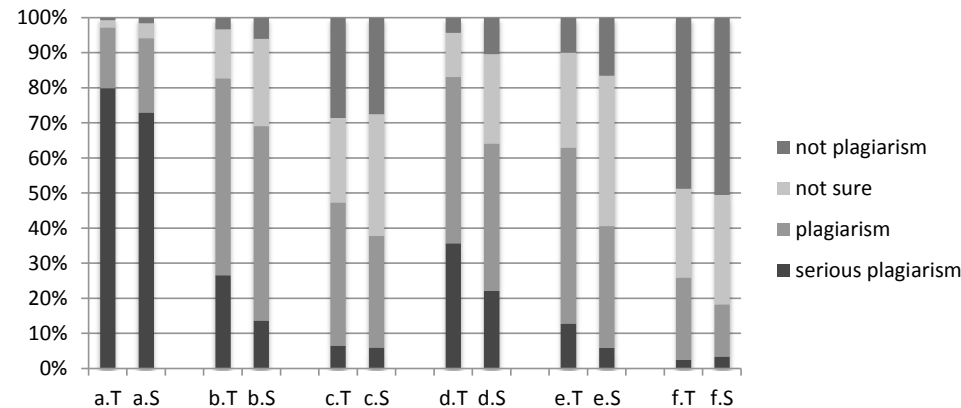


Figure 3. Teachers' and students' verdicts on plagiarism. Verdicts (serious plagiarism, plagiarism, not sure, not plagiarism) was requested for when 40% of an assignment is copied word for word and: (a) has no quotations; (b) has no quotations but has correct references without in-text citations; (c) as in (b) but with in-text citations; (d) has some words changed but no quotations, references or in-text citations; (e) has some words changed, no quotations or in-text citations but has references; (f) has some words changed, no quotations but has in-text citations and references

and students' perception of plagiarism, we have examined the verdicts (judgements: serious plagiarism, plagiarism, not sure, not plagiarism) given by students and teachers for a hypothetical case when 40% of a student's work is copied with varied use of citations and references (see a–f, in Materials and Methods). As shown in Fig. 3, there were differences in the percentage given by teachers and students for each type of verdict in each case (a–f).

More teachers than students gave a verdict of plagiarism when 40% of the assignment was copied and no references and in-text citations were used. The use of in-text citations and references appear to reduce both students' and to a lesser extend teachers' percentage of verdicts of plagiarism. It is also of interest to note that in all cases, a certain percentage of teachers and a much greater percentage of students were not sure or gave a judgement of no plagiarism. The percentage of such responses increased when references and in-text citations were used. Furthermore, when some words in the 40% copied part of the assignment were changed, the percentage of teachers and students who gave a verdict of plagiarism or serious plagiarism dropped. The above results indicate that both students and to a greater extend teachers are aware of plagiarism but are less likely to give a verdict of plagiarizing if quotations and/or references and in-text citation are used. This is most likely to be also the case, if some words in the part of the assignment copied have been changed as indicated here, which suggests that there is a need of a detailed definition of plagiarism that will not leave gaps in its perception by students or teachers.

Penalties for plagiarism

A major topic, which is related to students' and teachers' awareness and perception of plagiarism, is whether plagiarism should be penalized and what should such a penalty involve. Both, students and teachers were given a choice of 13 possible penalties and were asked to indicate the penalty known to them to be imposed if plagiarism were to be detected in a student's assignment or in a student's (undergraduate or graduate (master's)) final project (dissertation).

Analysis of the data showed that there were no differences in the% of teachers and students (range 49–50%) that chose a particular penalty to be imposed for plagiarizing in an assignment. The descending preference order of penalties chosen was the same for both teachers and students: “zero mark for the work” followed by “request to rewrite”, and then by “verbal warning”. There was less of an agreement between students and teachers responses for penalties given for plagiarism in a final project or dissertation. The penalty “zero mark for the work”, was the most popular choice for students whereas almost 50% of the teachers chose “request to rewrite”. More students (33 vs. 24%) chose “fail the module or subject” and this was the case for other forms of punishment like “expose student to school community” (17% vs. 8%), “suspend payment of student grant” (17% vs. 7%), “fail the whole program or degree” (31% vs. 22%) and “suspended from the institution” (29% vs 20%). The broader range of penalties chosen and the differences between students and teachers, most likely reflect the more lenient stance that teachers may take as a result of shared responsibilities in supervising student projects and having to consider the reputation of their Department/School of their Institution.

Policies and procedures

Both students and teachers were asked whether their Institutions had policies and procedures for dealing with plagiarism, about their availability and ways of administering these. Results were analysed in terms of “YES/NO” and “Not sure” answers and are shown in Table 2. As shown, more than 50% of students and teachers were aware of the existence of policy and procedures for dealing with plagiarism at their Institutions. As compared to teachers, however, a lot more students were not sure about them or how they were administered. These results suggest that HEIs in EU may need better ways of making faculty and students aware of policies/ procedures on plagiarism. One such way may be for teachers to undertake to communicate such information to students to the benefit of both.

Conclusion

This preliminary exploration of only a part of the data collected from the IPPHEAE survey have identified several interesting facts that may have an impact on plagiarism prevention policies and in designing anti-plagiarism courses for students:

- students are informed about plagiarism primarily from the web and not through lecturing or from teachers' guidance notes.

Table 2

Students' and teachers' views of policies and procedures on plagiarism, at their institutions

	Students			Teachers		
	YES	Not sure	NO	YES	Not sure	NO
This institution has policies and procedures for dealing with plagiarism	68%	24%	5%	75%	11%	13%
Plagiarism policies, procedures and penalties are available to students	54%	33%	10%	60%	27%	13%
Penalties for plagiarism are administered according to a standard formula	39%	49%	9%	50%	31%	18%
Student circumstances are taken into account when deciding penalties for plag.	22%	54%	20%	35%	45%	16%

- The most difficult aspect of academic writing for students is finding a good quality source for information retrieval rather than referencing.
- Most teachers misunderstand students' reasons for resorting to plagiarism.
- More teachers than students give a verdict of plagiarism for work that is plagiarized.
- Teachers' and students' opinions on plagiarism policies/procedures and on penalties do not agree.

Taking into account the above may help teachers to be more effective in trying to educate students about plagiarism, and HEI in implementing their policies on plagiarism by using a more fair approach in dealing and penalizing students who plagiarize. Considering that there were differences in the number of teacher and student participants between the various EU countries, with the highest number of responses coming from the Czech Republic, it is possible that the results obtained from this analysis, are biased. An analysis at national level may identify greater differences for some of the EU country participating. This is only a preliminary analysis of some of the data collected, and further comparisons using a country-by-country approach and even an Institution-by-Institution approach may help in using the observed differences for developing a more holistic approach in dealing with plagiarism in the different countries of the EU.

References

- Borg, E. (2009). Local Plagiarisms. *Assessment and Evaluation of Higher Education* 34 (4), 415–426
- Carroll, J. (2008). Assessment issues for international students and for teachers of international students, in *Enhancing internationalisation in business and management hospitality leisure sport tourism*, eds. Atfield, R. and Kemp, P., Newbury: Threshold Press
- DeVoss, D., Rosati, A. C. (2002). It wasn't me, was it? Plagiarism and the Web. *Computers and Composition*. Vol. 19, Issue 2, August 2002, p. 191–203
- Foltýnek, T., Čech, F. (2012). Attitude to plagiarism in different European countries. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*. 2012. vol. LX, No. 7, pp. 71–80. ISSN 1211-8516.

- Gu, Q., Brooks, J. (2008). Beyond the accusation of plagiarism. *System*, 36:3, 337–352
- Hayes, N., Introna, LD. (2005). Cultural Values, Plagiarism and Fairness: When plagiarism gets in the way of learning. *Journal of Ethics and Behaviour* 15 (3) 213–231
- McDowell, L., Brown, S. (1998). *Assessing students: cheating and plagiarism*. MARCET, University of Northumbria at Newcastle. ISBN 1861351151
- Pennycook, A. (1996). Borrowing Others' Words: Text, Ownership, Memory, and Plagiarism. *TESOL Quarterly*, 30: 201–230. DOI: 10.2307/3588141
- Risquez, A., O'Dwyer, M., Ledwith, A. (2013). "Thou shalt not plagiarise": from self-reported views to recognition and avoidance of plagiarism, *Assessment & Evaluation in Higher Education*, 38:1, 34–43
- Sutherland-Smith, W. (2005). Pandora's box: academic perceptions of student plagiarism in writing. *Journal of English for Academic Purposes*. Vol. 4, issue 1, January 2005, p. 83–95.
- Sutherland-Smith, W. (2008). *Plagiarism, the Internet, and student learning: improving academic integrity*. New York: Taylor and Francis. 2008. ISBN 0-203-92837-7

Acknowledgement

The IPPHEAE team is most grateful to EE2012 (Engineering Education 2012) and CEDE (Centre for Engineering and Design Education) for giving permission to adapt examples and templates on which the authors' guidance notes were based.

The research was funded by the project Impact of Policies for Plagiarism in Higher Education Across Europe (510321-LLP-1-2010-1-UK-ERASMUS-EMHE).

Authors are very grateful to all the project consortium partners for their cooperation in survey design and data collection: Coventry University (UK), Technical University of Łódź (Poland), Aleksandras Stulginkis University (Lithuania), University of Nicosia (Cyprus).

Authors

Tomáš Foltýnek, foltynek@pef.mendelu.cz,
Jiří Rybička, rybicka@mendelu.cz,
Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Czech Republic

Catherine Demoliou, demoliou.c@unic.ac.cy, Life and Health Sciences Department, School of Sciences, University of Nicosia, Cyprus

Copyright © 2013 Authors listed on page 127: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

ASSESSING THE VALUE OF A HOLISTIC USE OF TURNITIN TO PROMOTE ACADEMIC INTEGRITY

Adrian Lee, Antony Edwards

Abstract: This paper presents provisional findings of research evaluating Turnitin's use at a UK research intensive university. Its findings will inform institutional policy and practice and should also offer others an example of how the software can be used within their own contexts. The paper outlines staff use of Turnitin across the institution, reporting on an online survey, together with follow-up interviews. Findings indicate a divergence between departmental clusters in their use of Turnitin. Some clusters primarily use Turnitin to confirm and substantiate misconduct, in line with a motivation of detecting plagiarism, and another cluster uses Turnitin more formatively with students to discuss academic integrity and writing skills. Connected with these findings are results looking at why departments do not use Turnitin, and whether those that do offer students departmentally-based training in using the software formatively. The type of training staff received was found to have a positive impact on confidence with technical aspects of the software, and beliefs as to its impact on student awareness of plagiarism and quality of writing. Overall, this paper will contribute to the debate about the value of such tools within holistic academic skills and integrity processes.

Introduction

In recent years, many universities have become increasingly concerned to ensure the academic integrity of their students' work. At the University of York, this is achieved through a holistic approach involving staff and student bodies, in increased vigilance in assessment and marking, the investigation of cases of misconduct, educating students about the University's expectations and developing their core academic skills. Such a holistic approach has been widely advocated (Carroll and Appleton, 2001; Morris, 2010; Morris with Carroll, 2011; Sutherland-Smith, 2008). York is a research intensive institution of 15,000 students, 23% of whom are postgraduate and 24% are international students (University of York, 2012). Departments have considerable management autonomy, which brings challenges when seeking to communicate information, implement policy, and ensure a consistent understanding and application of key principles such as academic integrity. The Learning Enhancement Team (LET) works centrally to increase staff and students' understanding of academic integrity. The text-matching software Turnitin has been a part of this holistic approach since 2009, with an emphasis on formative learning and students' responsibility for their work, as well as more disciplinary applications of the program.

This paper introduces an impact evaluation of the University's deployment of Turnitin, discussing preliminary findings from an ongoing study. It first elaborates on the institutional context and explains how Turnitin is used. Second, it outlines the methods used which include staff surveys and interviews. These have captured how and why Turnitin is used and why it might not be used, perceptions of its purpose and effect on academic integrity, students' writing, and on staff working practices. Results will then be presented and emerging key themes discussed. These themes

include the need for well-publicised information about how Turnitin can be used, the need for appropriate user training and for Turnitin to be efficiently integrated into the assessment, marking and feedback systems. The paper concludes by outlining how the evaluation will be completed and the findings utilised to ensure the University most efficiently and effectively uses Turnitin within its holistic culture of academic integrity.

The institutional context

In common with many UK universities, York has recently been expanding and diversifying its student body to include students potentially less familiar with UK higher education. This has created a need to explain academic writing practices clearly. Together with other academic integrity resources, all students have been given access to Turnitin to promote effective source use and to develop their writing skills. To situate the findings of this study, we first present the institutional context.

The University is divided into 30 autonomous departments, each with a direct relationship with senior management, and no intermediate faculty structure (University of York 2013). In this paper, departments have been grouped into the three cluster types: Arts and Humanities (Arts), Social Sciences, and Sciences. Within each cluster, individual departments vary in size, the ratios of undergraduates to postgraduates, and student demographics (University of York, 2012). In 2008/09, the University purchased a Turnitin licence for originality checking, with a focus on using the tool formatively with students to complement their reading, source use and writing skills (Wiggins 2010). Since 2009, the LET has delivered training workshops for students, which introduce the purpose of Turnitin, instruct them on using it through a Virtual Learning Environment (Blackboard) and to interpret originality reports. The workshops aim to enable students to assess the integrity of their work, and to give them an opportunity to redraft and submit assignments meeting University requirements.

Two types of workshops are offered: departmental workshops for year groups/ whole cohorts, and central inter-departmental open workshops. Both are organised centrally by the LET, and delivered by its staff and a team of “postgraduates who teach” (PGWT). Various authors have proffered the need for such active, participatory training to crystallise for students the concept of plagiarism and to develop confidence using Turnitin (Dahl, 2007; McCarthy and Rogerson, 2009; Rolfe, 2011). Our approach to student training, however, contrasts with smaller-scale, individual faculty-led or academic-led formative models used elsewhere (Badge and Scott 2009; Hunter 2012; Rolfe 2011; Starr and Graham-Matheson 2011). There is no University-wide policy governing the use of Turnitin, instead departments decide whether to organise workshops, with the LET working closely with departments to meet any particular needs.

The varied departmental cultures and autonomy has led to considerable differences in the use of and attitudes towards Turnitin. Individual academics use Turnitin in different ways, with the decision whether to use it to assess originality sometimes left to individual academic judgement. There are no formal reporting procedures as to its use, in what way and for what purpose. This has implications when supporting users, who might not know what help is available, but also from the University’s point of view in ensuring Turnitin is used appropriately and transparently. As institutional experience

of using Turnitin grows, it is pertinent to explore how staff and students use it, why and with what perceived impact.

Empirical Methods

The study began in Autumn 2012 with staff-focussed research. An online survey was devised using Google Forms (Google, n.d.). Adapted from Starr and Graham-Matheson's (2011) survey, it asked staff, including PGWT, about themselves, their personal knowledge of Turnitin, their use of it, their knowledge of its use by colleagues and their perceptions of its impact on working practice, student behaviour and learning. A pilot was undertaken with a small group of colleagues that had previously spoken to the LET about Turnitin. Based on the pilot responses, small changes were made to the survey.

Members of staff were asked to indicate their willingness to take part in an interview to discuss their perspectives and experiences. From those willing to be interviewed, a representative sample was contacted based on cluster and their reported use or non-use of Turnitin. Staff were interviewed using a semi-structured format, following a topic guide relevant to their use of Turnitin. The aim was to add detail to the survey data, and to explore related issues around holistic approaches to promoting academic integrity, detecting and investigating misconduct and the use of text-matching tools generally.

The second part of the research, not reported in this paper, has been a similar survey of students. The next stages of the study will involve follow-up interviews with students and an assessment of the institutional database of Turnitin originality reports. We plan to examine assignments students upload when drafting work, how often they submit drafts of the same assignment, what matches Turnitin reports highlight and how students appear to revise work in light of these matches.

Research Results

This section summarises key findings from the staff survey and incorporates data from the 19 interviews (17 academics and two in support/ administrative roles). In total, 130 staff completed the survey: 90 Academics, 22 in administration or support roles and 18 PGWT. Figure 1 illustrates the academic clusters of participants' departments, showing that 40 respondents currently use Turnitin, 23 have used it previously but do not currently, and 67 have never used it as a member of staff. It also separates staff into their subject clusters: 33 from the Arts, 49 from Social Sciences and 48 from Sciences. The following discussion of results considers these patterns of use in conjunction with Tables 1 and 2, which respectively present why and how current users reported they used Turnitin, and Table 3 which asked non-users and non-current users why they did not use it.

The first theme that emerged is a split between using Turnitin as a plagiarism detection tool and as a formative tool. Table 1 shows greater use in the Social Sciences and Sciences to detect plagiarism than in the Arts where use to help students understand what plagiarism is and to improve their writing predominated. This difference between clusters is also echoed in Table 2, which shows the Arts using Turnitin as a developmental tool and colleagues in the other clusters employing it to confirm and

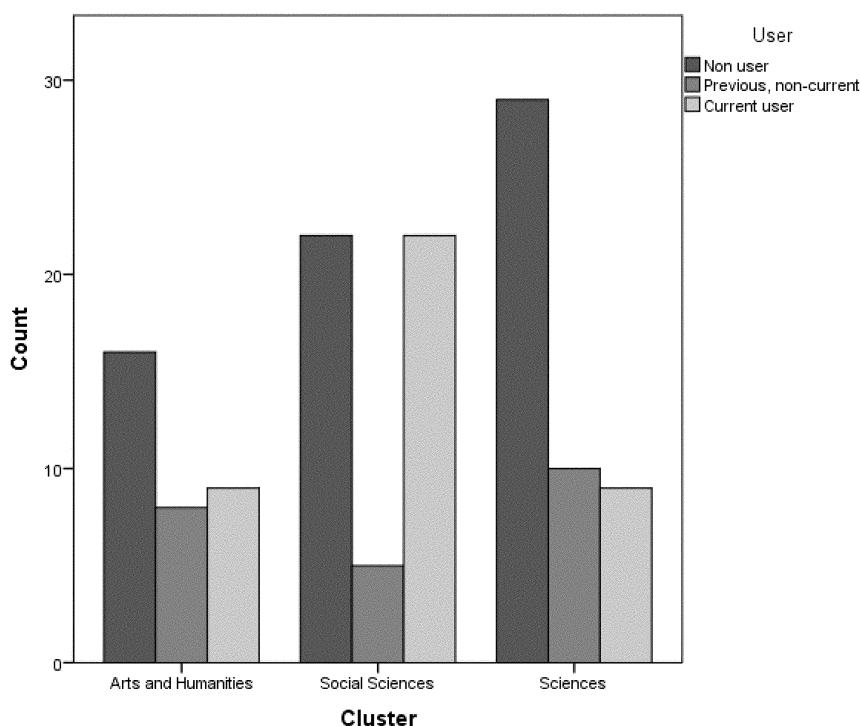


Figure 1. Respondents by academic cluster and reported use of Turnitin ($n = 130$)

substantiate misconduct, in line with their key motivation of detecting plagiarism. The distinctions were evident in the interviews, but they also suggest greater complexity:

So we use Turnitin only when we suspect something is plagiarised, so we don't put everything through Turnitin. There are some cases, there's one module in particular which uses it as a formative tool, so students are asked to submit their work through Turnitin as a formative exercise and feedback is then provided based on that Turnitin report. (Participant 3, Social Sciences).

To improve academic writing right from the start, as early as possible, and because these first [Masters] essays they're pretty hard to read. Not all of them, but usually they are... And of course it's just simple plagiarism detection. And possibly compiling the evidence in a very effective efficient way. (Participant 7, Science).

In terms of non-users and previous non-current users, we felt it useful to establish why Turnitin is not employed. It was speculated that this might be due to perceived weaknesses in the software or lack of knowledge of its application, in line with the above distinction between formative and detective uses. Responses to the listed options in the survey, and others that participants volunteered, are in Table 3.

In Table 3, the most cited reason for not using Turnitin was that "assessments in my modules do not necessitate using Turnitin", 34% overall, with this being particularly

Table 1

Why “current users” used Turnitin

	Arts and Humanities <i>n</i> = 9	Social Sciences <i>n</i> = 22	Sciences <i>n</i> = 9	Total <i>n</i> = 40
To detect plagiarism	22%	82%	78%	68%
To help students understand what plagiarism is and how to avoid it	67%	36%	33%	43%
To deter students from plagiarising	33%	41%	44%	40%
To encourage/help students to improve their academic writing skills	56%	27%	33%	35%

NB: In this (and subsequent) tables participants could give select multiple responses, so percentages may sum over 100%.

Table 2

How “current users” report using Turnitin

	Arts and Humanities <i>n</i> = 9	Social Sciences <i>n</i> = 22	Sciences <i>n</i> = 9	Total <i>n</i> = 40
I used Turnitin to check the originality only of work I suspected of containing unacknowledged source material	22%	68%	44%	53%
I used Turnitin formatively with students to discuss academic integrity and writing skills	67%	27%	22%	35%
I used Turnitin to check the originality of ALL my students’ assignments	0%	23%	22%	18%
I gave students the opportunity to view the originality report for their final submission	0%	14%	11%	10%

given by Science participants (41%), whose colleagues we previously noted were using the software for detection purposes. However, another highly selected option was “I do not know enough about how Turnitin can be used to assess originality” (26% overall) with a further 7% indicating they did not know enough generally about what Turnitin is and how staff can use it. Such points were expanded upon when interviewing staff:

The problem is it’s a thing that’s going to take more of my time, or so I perceive it. And I don’t have that much time... I don’t perhaps understand, I mean, having talked to my colleagues I think probably I do need to go and try it out for myself, and get a clearer idea, I’ve always intended to do that, it has just not been high enough up my list of priorities. (Participant 4, Arts).

These [tasks] are not in English, these tend to be formatting texts if they’re typing things up: mathematics, computer programs... For summative assessments, at the moment they are entirely in an examination hall with a pen and paper . . . Because for most of us,

Table 3

Reasons why participants did not currently use Turnitin

	Arts and Humanities <i>n</i> = 24	Social Sciences <i>n</i> = 27	Sciences <i>n</i> = 39	Total <i>n</i> = 90
Assessments in my modules do not necessitate using Turnitin	38%	22%	41%	34%
I do not know enough about how Turnitin can be used to assess originality	13%	30%	31%	26%
I am happy with the existing strategies and tools I use to identify plagiarism (eg. Google searches, SafeAssign, changes in writing style)	8%	15%	18%	14%
Other reasons	13%	15%	5%	10%
It is not part of my role	13%	7%	5%	8%
I lack awareness about Turnitin overall	4%	4%	10%	7%
My department has agreed not to use Turnitin	4%	0%	8%	4%
I do not have electronic copies of assignments	0%	4%	5%	3%
I view Turnitin as ineffective in detecting misconduct	0%	4%	5%	3%

we're not asking students to write English, or natural language, the only place where it really happens is in the ISM [Independent Study Module]. (Participant 6, Science).

The individual respondents' reported use, non-use and explanations thereof were compared with their responses indicating if they knew who else in their department uses Turnitin. Results indicated that many participants were unaware of use (or lack of) in their department: 85% of Arts, 51% of Social Sciences, and 54% of Science. However, this is not the case for every department, as the survey and interviews identified departments where use, and reports on colleagues' use was reported as known. In the interviews this appeared to relate to the administrative roles several participants undertook. The quotes below highlight typical levels of awareness from Turnitin users and non-users:

We had a discussion about this at Board of Studies at the start of this year . . . a fairly lengthy discussion that one of our experimental officers did a presentation on it to bring a lot of the staff up to speed with what we do and why we do it, really, and to, I think really, to raise awareness of it, to see if there's more we could do in the future, to remind people that it is there on the VLE, that they can use it if they want to. (Participant 15, Arts).

I'm not the best attender at Board of Studies I'm afraid . . . I will attend Graduate School Board, and we have had discussions about Turnitin, but I'm not really, I can't remember them sorry. I can't remember. I mean, are we supposed to be using it yet? Are we all supposed to be using it now? (Participant 10, Social Science).

Table 4

Perceived knowledge of how Turnitin is used within respondents' departments

	Arts and Humanities <i>n</i> = 4	Social Sciences <i>n</i> = 24	Sciences <i>n</i> = 21	Total <i>n</i> = 49
To check work that, on marking, has raised suspicions of academic misconduct	50%	63%	24%	45%
To check assignments on individual programmes/ modules only	0%	17%	43%	27%
To strengthen the evidence in an academic misconduct case	25%	33%	24%	29%
I do not know specifically, but I know colleagues do use it	25%	13%	29%	20%
To routinely check ALL assignments students submit for possible academic misconduct	0%	21%	5%	20%
Other	25%	8%	5%	8%

Staff usually reported that their department's use was similar to their personal use (see Table 4). 45% of staff indicated that Turnitin is used after initial marking, rather than being an integral part of the marking process, with the majority of the Social Science respondents selecting this option (63%). Such participants are using Turnitin to check any concerns, possibly providing evidence for a misconduct case, after first using their academic judgement and their ability to first spot potential plagiarism. Table 4 also shows that, in the Sciences, Turnitin is used more on individual assignments than a widespread basis. However, as with other results we have reported, our data illustrate different marking practices, several staff routinely check the similarity index summary view before marking to see if any assignments stand out as potentially problematic:

I look for similarity... I start by looking at what sort of percentages we are talking about. Let's have a look at say, this one, 68%. Why have we got such a big similarity? (Participant I13, Social Science).

When I'm marking their work, that's when I normally run across Turnitin and it's usually when it's applied to larger essays, so the third year essays and the placement... essays that the students write, those are the things that are normally passed through Turnitin and are then handed off to us as markers, as part of the packet of the student work. (Participant I17, Science).

It was also reported that some departments' assessment processes mean administrators produce originality reports and highlight those with higher similarity indexes to markers, for example:

I'm running the stuff through Turnitin, so all of the assessments, the essays that we get... I check the initial reports that we get... and then I send out emails and say "this many of your students have got over 25%. Go and have a look and please do have an overall look as well". (Participant 16, Social Science).

Table 5

Why participants believe departments offer student Turnitin workshops

	Arts and Humanities <i>n</i> = 4	Social Sciences <i>n</i> = 24	Sciences <i>n</i> = 21	Total <i>n</i> = 49
To raise the students' awareness of what plagiarism is and how to avoid it	15%	11%	24%	84%
To develop the students' academic writing skills	77%	79%	41%	65%
To provide students with an opportunity to check their work formatively before the department checks it summatively	46%	68%	53%	57%
To develop the students' referencing skills	62%	63%	41%	55%
To counter student anxiety about plagiarism	38%	53%	29%	41%
To meet students' demands for skills support/access to learning resources	38%	53%	24%	39%
I do not know	8%	11%	12%	10%

Staff were also asked whether or not their department organised training workshops for students. The results suggest many did not know whether workshops were offered (40% in Arts, 47% in Social Sciences and 44% in Sciences, the majority of these being academic members of staff). Moreover, when staff selected that their department did or did not organise workshops, we saw responses of both “Yes” and “No” from the same department in five cases, and “No” from respondents in departments where workshops were indeed delivered. We can infer a lack of communication between staff, and possibly miscommunication from them to students about these matters. When asked why departments did not organise workshops 12 of the 24 respondents did not know and 12 offered a variety of responses. The most cited were: “Because we feel students should already know what is expected of them” (4), “Because we collectively agreed that student use of Turnitin encourages bad academic practice” (3) and “Because we don’t know enough about how Turnitin can be used to develop students’ academic skills” (3).

The 49 staff who did indicate that their department offered training to students were also asked why (see Table 5). Note that the LET “markets” workshops as primarily a tool for improving the integration of source material and an aid to writing skills, and then to prevent and counter plagiarism. Recall also from Table 2 that 67% of Arts, 27% of Social Science and 22% of Science respondents indicated using Turnitin as a formative tool to promote student discussion of academic integrity and writing skills. In the Arts and Social Sciences the most common reasons for delivering training were developing writing and referencing skills, with the Social Sciences also identifying that it gave students the chance to formatively check their work. In the Sciences, formative checking was the most commonly-selected reason, which is surprising given few staff claimed to use Turnitin for this purpose with their students.

Survey questions asked staff to comment on the impact they perceived Turnitin has had on staff and student awareness of plagiarism, students' writing and assessment and misconduct case practices. Here, impact was measured on a scale of "positive" (3), "no impact" (2) or "negative" (1). T-tests showed that staff who said their department organised student workshops perceived a significantly more positive impact on students' awareness of plagiarism (mean difference = +0.542, SE = 1.83, $p < 0.01$) and on the quality of students' writing (mean difference = +0.383, SE = 0.148, $p < 0.01$) when compared with those that did not offer training. The interviews highlighted the speculative element of what impact Turnitin has had, for example:

I think that, delivered very early on, as in, first term, a tutorial which introduces them to this tool and perhaps gets them to put a piece of work through it, on a completely formative basis, might be helpful in demonstrating to those students who still don't really understand what plagiarism is, and how it can affect their own writing, it could be useful in demonstrating to those students, "well this is what it is, and this is how it can affect you, and this is how you can remedy it". (Participant 11, Arts).

The conversations you have with students when you're in the workshop with them. Those sort of, the light bulb moment shall we say. Those where they see actually, although they might have got A's in their A level results and their A level work, and they put that through Turnitin and they realise "oh actually, hang on a minute, there's a problem here. Now we're on a different level". That light bulb moment is the impact I see. (Participant 3, Social Science).

However, in one case the benefits were more tangible:

We had tutorials, I think a week before submission... for this Masters module, where students were shown how Turnitin works... And I could see the difference in the submitted coursework, just by reading it... I looked through the material without using Turnitin, marked it, and I could already see that this was much smoother at least. (Participant 7, Science).

The survey then looked at the training and support given to staff users, and about their confidence using the software (see Figure 2). They were asked if they were "very confident", "confident", "not very confident" or "not confident at all" to (a) add submission points to a Blackboard module site, (b) upload documents to Turnitin and access the Originality Reports, and (c) interpret the Originality Reports. The survey also recorded the source of training or support and the results were coded into those that indicated they had received it from the centrally-based LET or E-Learning Team (who are responsible for Blackboard) or not. This is to give a sense of whether staff were supported by the "official" central service or another, such as informal peer instruction. Of the 52 users who indicated they had received some formal or informal training or support, 35 had received "official" training. We compared the confidence of those with and without training (measured with a "4-3-2-1" scale), using a MANOVA, finding a significant effect of training on the groups' confidence, $F(3, 48) = 0.181$, $p < 0.05$ (Pillai's trace). There were significant between-subject effects in terms of (a) setting up submission points, with an average of 0.64 confidence increase ($p < 0.05$), and (b) uploading files and accessing originality reports ($p < 0.05$), with an average of 0.71. There was no significant difference in confidence in (c) interpreting originality reports.

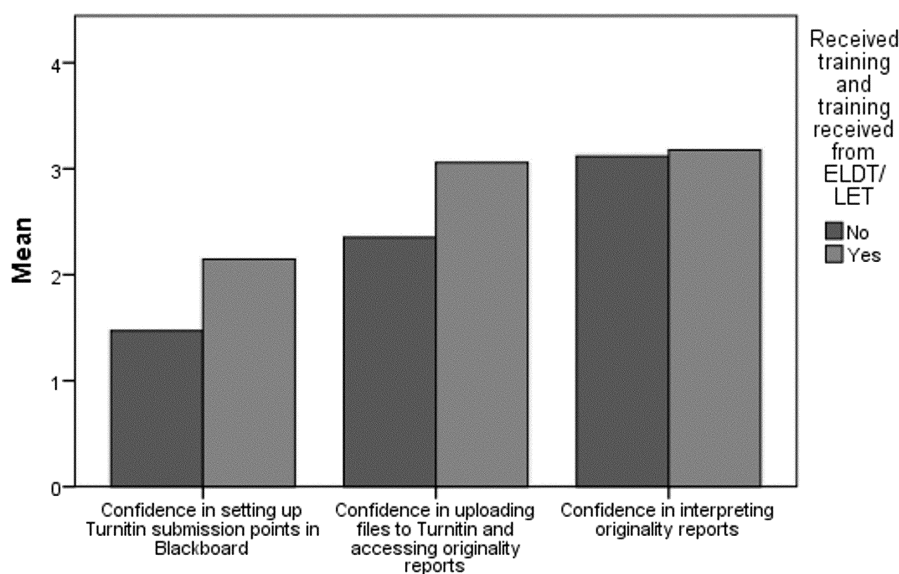


Figure 2. Mean confidence in Turnitin use for staff who had received central and non-central training ($n = 52$)

We did not find that length of use (more than three years, one-to-three years, and less than one year) was a significant factor when we included it in the model as a covariate ($p > 0.2$).

The survey and staff interviews have provided a useful snapshot of how Turnitin is perceived at the University. Considering the results in the institutional context suggests a number of key themes for further exploration to confirm Turnitin's benefit and how best to provide staff and student access.

Discussion

This study sought to understand to what extent staff use Turnitin, why they use or do not use it, their perceptions of use and the impact they perceive it has on academic integrity, writing skills and assessment procedures. We now reflect on the above results, considering the role of the department and the individual on Turnitin use.

The outcomes of how and why respondents use Turnitin (Tables 1 and 2) differed by cluster. In the Arts, nine people said they currently used Turnitin. A department that was highly represented in this cluster reported using Turnitin to help students understand what plagiarism is and how to avoid it, encouraging them to improve their writing skills, over detecting plagiarism. In the Social Sciences the overriding reason for using Turnitin was to detect plagiarism (21 people), with two departments accounting for the majority of these responses. The main reported use (checking work that on marking had been suspected of misconduct), saw eight of the 27 people coming from one social science department alone, though three of these respondents also said they

used it formatively. This is noteworthy as the department seemed to show a quite uniform pattern of use and knowledge of how their department used Turnitin. Such intra-departmental similarity was not widespread in other departments that had larger numbers of respondents.

Reasons that emerged for non-use indicate a need for more available information and staff training. Science colleagues felt that Turnitin would not be particularly effective or was unnecessary in monitoring integrity, due to the factual and descriptive nature of lab report write ups, as well as a predominance of computing languages in assessments. However, the interviews suggested that assignments on some courses could be suitably assessed by Turnitin. Some Arts survey responses also indicated assessments do not necessitate using Turnitin (38%), but members of this cluster tended to still promote formative aspects of Turnitin access. The expectation here is that such subjects use traditional written tasks, expect critical discussion and the use of source material, thus the result stands out. The majority of such responses were from an Arts department whose students produce practical, non-essay based assignments more similar to the Sciences. It is important therefore to note that not all types of student assessment might benefit from text-matching analysis, and that Turnitin might not be the most suitable tool for this. There also seemed to be a lack of openness in some departments, to considering Turnitin's potential value alongside other strategies for assessing academic integrity that were more highly favoured. Furthermore, the data for one Arts department showed a collective perception that staff were not permitted to use Turnitin for "policing" integrity, instead students could only use it for formative purposes. It is proffered that this is due to ambiguous initial briefings and a lack of later engagement in training.

Many staff appeared unaware that students had access to training, suggesting a need to better promote the student workshops to departments. Some were unsure of whether workshops were organised by their departments, or were incorrect in believing that none were offered. Of the 24 academics who believed their department did not offer student workshops, the first author personally organised sessions for seven departments in which 18 of these academics work. Such academics will lecture on programmes and/ or act as students' personal supervisors, and arguably, they should be expected to know the resources provided as part of their departmental study skills training.

The survey participants' awareness of how Turnitin can be used, and for what purpose, implies a need for more conveniently accessible staff training. It is encouraging that the survey suggests that those who had received central training had a more positive view of the value of student access, and reported more confidence in using the software themselves. Though a programme of staff training already exists, this study highlights a need to develop and deliver it in ways that inform and empower staff to use Turnitin effectively, fairly and transparently in assessing students' academic integrity.

Conclusion

This study aimed to scope out how widely Turnitin is used across the University of York by staff and students and to understand how users view its impact on academic

integrity, writing and assessment processes. This paper has presented the results of the staff survey and interviews. Emerging themes point to a diverse use between clusters, indicating a need to provide more information to staff about how they can effectively and efficiently use Turnitin, as well as what potential benefits there are for encouraging students' formative use. The student survey will hopefully provide data about how students use Turnitin and what value they believe it has for their attainment, which can be used to inform staff with currently limited awareness of impact, and in some cases concern about students inappropriately using the software when writing. Thus far, the study has proven insightful for the Learning Enhancement Team in demonstrating how and why different academic departments use Turnitin and indicating where we can inform staff of its potential. We aim to create a holistic culture of academic integrity that seeks to educate members of the University community as well as promoting vigilance to misconduct and systematic, appropriate handling of cases. For this to happen, we will need staff and student support and cooperation, in the context of time-pressured academic workloads. As the study continues and reaches its conclusion, it is expected that our findings will enable the LET to lead the University's initiative to achieve such goals.

References

- Badge, J. & Scott, J. (2009) *Dealing with plagiarism in the digital age*. Leicester: University of Leicester.
- Carroll, J. & Appleton, J. (2001) *Plagiarism: a good practice guide*. Oxford: Oxford Brookes University/JISC.
- Dahl, S. (2007) Turnitin: the student perspective on using plagiarism detection software. *Active Learning in Higher Education*, 8 (2), 173–191.
- Google. (n.d.) Google Apps for Education. Available from <http://www.google.com/enterprise/apps/education/> [accessed 28 February 2013].
- Hunter, A. (2012) "Text comparison software for students": an educational development tool or quick "text checker"—examining student use and perceptions of value. Presented at 5th International Plagiarism Conference, 16–18 July 2012, Gateshead. Available from <http://www.plagiarismadvice.org/research-papers/item/text-comparison-software-for-students-an-educational-development-tool-or-quick-text-checker-examining-student-use-and-perceptions-of-val> [accessed 28 February 2013].
- McCarthy, G. & Rogerson, A. (2009) Links are not enough: using originality reports to improve academic standards, compliance and learning outcomes among postgraduate students. *International Journal for Academic Integrity*, 5 (2), 47–57.
- Morris, E. (2010) *Supporting academic integrity: approaches and resources for higher education*. York: Higher Education Academy.
- Morris, E. with Carroll, J. (2011) *Policy works: recommendations for reviewing policy to manage unacceptable academic practice in higher education*. York: Higher Education Academy.
- Rolfe, V. (2011) Can Turnitin be used to provide instant formative feedback? *British Journal of Educational Technology*, 42 (4), 701–710.
- Starr, S. & Graham-Matheson, L. (2011) Efficacy of Turnitin in support of an institutional plagiarism policy. A report for plagiarismadvice.org on research conducted at Canterbury Christ Church University. Available from http://www.plagiarismadvice.co.uk/images/documents/starr_04.pdf [accessed 28 February 2013].

- Sutherland-Smith, W. (2008) *Plagiarism, the internet and student learning: improving academic integrity*. London: Routledge.
- University of York. (2012) University of York—annual student numbers (FTE). Available from <http://www.york.ac.uk/admin/po/statistics.htm> [accessed 28 February 2013].
- University of York. (2013) Academic departments, research and teaching centres. Available from <http://www.york.ac.uk/about/departments/academic/> [accessed 28 February 2013].
- Wiggins, C. (2010) Turning Points: building a framework for active student engagement and learning with Turnitin. Presented at 4th International Plagiarism Conference, 21–23 June 2010, Newcastle-upon-Tyne. Available from <http://archive.plagiarismadvice.org/conference/previous-plagiarism-conferences/4th-plagiarism-conference-2010> [accessed 28 February 2013].

Acknowledgements

The authors acknowledge the advice and information given by Simon Starr, Canterbury Christchurch University, which assisted the design of the surveys used in this study.

Authors

Adrian Lee, adrian.lee@york.ac.uk,

Antony Edwards, ant.edwards@york.ac.uk,

University of York, United Kingdom

Copyright © 2013 Authors listed on page 136: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

ELECTRONIC PLAGIARISM DETECTION SOFTWARE AS SELF-TEACHING TOOL FOR PLAGIARISM AVOIDANCE IN BIOSCIENCE UNDERGRADUATES

Shiva D. Sivasubramaniam

Abstract:

Following the initial enhancement of plagiarism detection by Turnitin®, academics started to use it successfully. However, many academics are using it as detection software, without using this as a tool to enhance the students' learning. This study was designed to evaluate the potentials of Turnitin® software for educating students to avoid plagiarism. This study was carried out with two cohorts of level 1 undergraduates from 2008 to 2009. After the initial induction about good academic practice, students were set an academic essay. They were allowed to use the Turnitin® software only once to improve their essays, before final submission. After their re-submission, the articles were again checked for plagiarism using the software and by proof reading in comparison with original articles.

The results from their first attempt have shown that 37 out of 70 students (approximately 53%) were able to write a scientific essay with proper attributions and referencing. Of the remainder, around 11% (8 out of 70) of the student submissions had worrying amounts of plagiarism and the rest had incidences of inappropriate paraphrasing (5 students 7%), copy-pasting without proper referencing (5 students 7%) and "patch-working" (15 students 22%). Interestingly, the data from the second submissions showed that at least 59% of students (22 out of 37) were able to re-write their essays without plagiarising. The rest were shown improvements in their academic writing. The data from this study suggests that it is possible to use Turnitin® as a learning tool to improve the students' academic writing.

Background—Historical prospective:

Many "text-matching" tools have been developed to detect inappropriate use of information and plagiarism. However the introduction of Turnitin® software has revolutionised the detection process. Since its introduction, several academics have reported that compared to other tools, this software is superior as it is simple to use, accurate in matching the broadest range of sources, embedded with grade marking facilities and above all has the potential to be used as an educational tool (Sivasubramaniam, 2002; Satterwhite and Gerein 2003; Gauder, 2004; Jocoy and DiBiase, 2006). It is an accepted fact that Turnitin® has tremendously reduced the time spent by the academics when checking work for plagiarism. Jocoy and DiBiase (2006) reported that it detected five times more instances of plagiarism than manual methods. Moreover, by highlighting where malpractice has occurred it provides definitive proof for academic misconduct inquires. However as Wright and Nigel (2008) correctly pointed, the software "should be regarded only as a tool to assist in the detection of plagiarism". The actual decision making (whether it is plagiarism or not) actually relies on academic wisdom.

Others, like Jones (2008) further emphasized that Turnitin® is not a plagiarism detection service but simply a text-matching system a fact accepted by many experienced

academics who use this matching to decide academic integrity. Thus like any other e-learning devices, Turnitin® only assists, rather than replaces, human decision making. More importantly it is clear that increased use of this software has in itself resulted in plagiarism deterrence (Sivasubramaniam, 2004, 2006)

After its initial usage for “plagiarism detection”, Turnitin® was transformed into a grade marking tool. This has opened up the opportunity of using this software for teaching and/or formative feedback. Even during its introductory phase, reports claimed that students would like to use this tool to check their work before final submission, since they would be able to learn from their mistakes and improve writing skills (Sheridan and Brake, 2001; Dahl, 2007). In fact Turnitin® reports themselves can be used as feedback. Therefore it is clear that using Turnitin® feedback together with redesigning of the assessment can encourage students to be interactive with academics (Nicol and Draper 2008). Other authors went further and proposed using it as a teaching tool to increase (a) the understanding of plagiarism (b) its avoidance and (c) competencies amongst students (O’Hara et al 2007; Cheah and Bretag, 2008). Thus in 2009, Davis and Carroll concluded that Turnitin® can be used as a formative assessment tool to help students avoid plagiarism and improve citation practices.

Introduction

Plagiarism amongst undergraduates has been a well-studied and researched concept since the creation of the Plagiarism Advisory Service (JISC-PAS) in 2002. JISC-PAS has not only highlighted plagiarism practices amongst undergraduates but also assisted UK academic institutions to use Turnitin® software for free. The publication of Turnitin® and other detection software’s, made the “detection” quicker and also played a large part in reducing plagiarism in universities (Wicker, 2007). Nottingham Trent (NTU) was one of the first universities to adopt electronic detection since its introduction in 2002. Interestingly in Biomedical Science at NTU, this increased the detection rate from 4 per year (average) before 2002 to 18 per year in 2003 (Sivasubramaniam, 2004; 2006). This is largely due to better and effective detection by this software. As a result of the improved detection rate, several academics within NTU started to use it; some used it only to confirm suspected plagiarism and others used it as a routine for all coursework. Some others even made the reports available to the student viewing. However, the School has a strict policy of not allowing the students to use this software for student viewing and re-submission. It has been argued that the student would use this software simply to reduce the “high scoring” and getting away with plagiarism, just by quoting, patchworking, rearranging the sentences and referencing (Australian universities teaching committee, 2007).

Other educational organisations have successfully been using this software as a tool for educating students. In fact Becton (2007) has pointed out that the software can be a useful tool for teaching students about plagiarism and assisting them to avoid it. By highlighting the origin of a “patchwork”, it would show the original reference and reinforce the need for appropriate citation. However this is only possible if the report is well interpreted by the students. Thus, it is still not clear whether this software can be used as self-studying tool for improving academic writing by “learning from

their mistakes". Carroll and Jameson (2009) have correctly pointed out that Turnitin® cannot be a substitute for all the careful teaching and feedback. In fact it is generally believed that academic input, in the form of tutoring/mentoring (spending some time with the students) to explain the originality report, is necessary. This pilot study aimed to investigate whether the Turnitin® software can be used as a self-studying tool to improve the academic writing of level 1 undergraduates.

Methods

The study was carried out in the School of Science and Technology, Nottingham Trent University; Nottingham UK Level 1 undergraduates from the 2008 to 2009 cohorts were used in this study as a part of the workshop on how to avoid plagiarism. In the workshops, the students were first introduced to good academic writing and explained the ways to avoid plagiarisms. They were also trained to use the Turnitin® software. Then a task in a form of a scientific essay was set; each student was given a separate scientific title.

They were told to submit their essays to the software, read the originality reports and correct any "potential" plagiarisms and re-submit for academic assessment. Analysis of student learning was carried out by comparing their first and re-submitted reports.

The following criteria were used to check the authenticity of student learning:

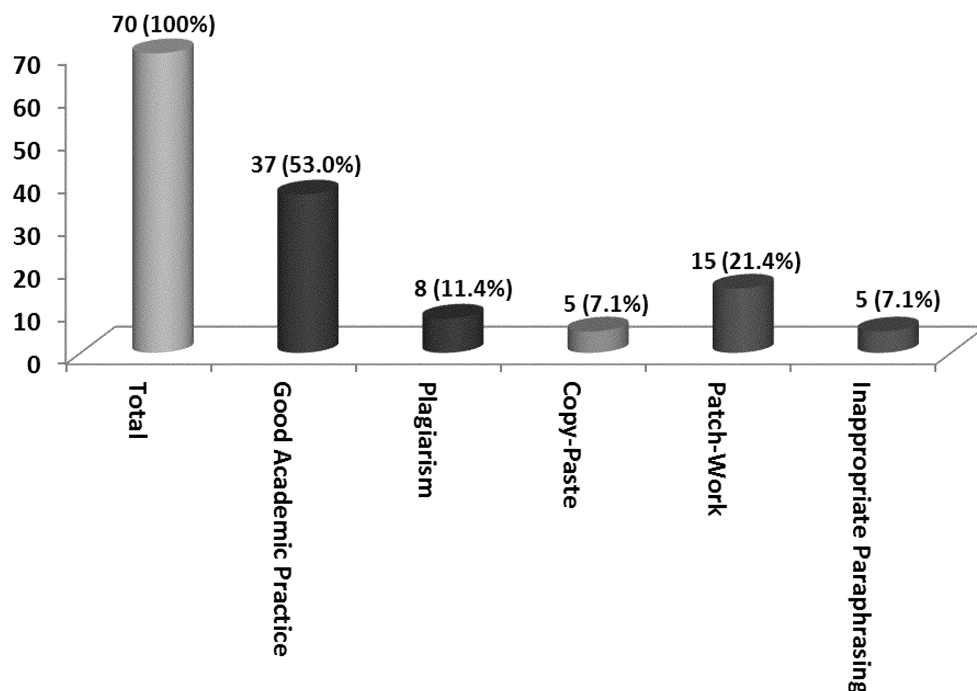
1. Number of student reports with good academic writing
2. Reports with "patchwork"
3. Copy-pasting with proper referencing
4. Inappropriate paraphrasing/summarising.

Although the similarity index provided by the software was used as guide, the final decision for categorising into above criteria was made by the academic.

Results

The results from the first and the re-submissions are summarised in figures 1 and 2. From figure 1, it was interesting to note that 37 out of 70 students (approximately 53%) were able to write well-referenced essays with appropriate citations. Although the academic rigour of these essays varied (marks ranging from 45% to 89%), these students were able to read, understand the concept of their title and then write a well attributed essays. On the other hand the rest of the essays (around 47%) had significant incidents of "bad academic practice" or were slightly plagiarised. However, only 8 out of 70 student (around 11%) essays were blatantly plagiarised. Many of them (15 students) used patchwork to bring the essence into their essays. While other essays had (a) incidences of inappropriate paraphrasing/summarising (5 students); or (b) copy-pasted sentences without proper attributions (see figure 1).

It was interesting to note that out of 33 final re-submitted essays 19 students (approximately 58%) managed to re-write (or amend) their essays into acceptable submissions. However, most of these students either used attributions using quotes



Bars represent actual numbers with calculated percentages in brackets. Good academic writing refers to the essays without (or with negligible amounts of) bad academic practice; copy-paste refers to the essays with sentences lifted from the original articles without proper attributions

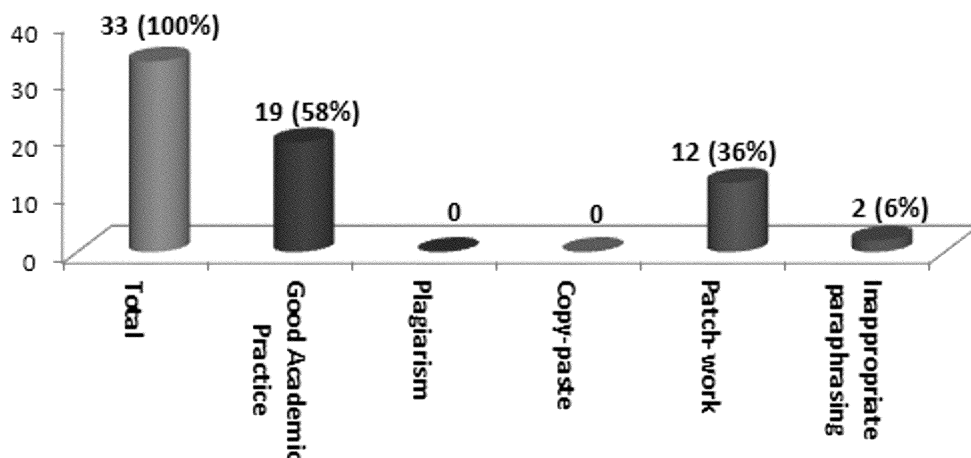
Figure 1. Student performance—First submission

and citations or tried to paraphrase or summarise the original sentences with proper referencing. Almost all of these students produced a good reference list.

Another 12 student essays (36%) showed incidences of patch working. In other words students tried re-word the original sentences just by changing the format and/or grammar. Finally very few students [around 2 (6%)] tried to paraphrase the original sentence without any success (see figure 2). From this the author was able to differentiate the students who could improve their academic writing by self-learning from the Turnitin® similarity reports (12 students) from the ones who would need academic assistance or tutoring to improve their academic writing skills.

Interestingly, during the tutorial session, the author was able to identify certain common misconceptions about “good scientific writing” amongst the students. These are summarised in table 1.

Although it is difficult to show entire reports, example paragraphs from the first and re-submissions are compared in table 2 to show how students have improved their reports using Turnitin® feedback.



Total numbers of re-submitted articles were 33; Note the incidences of plagiarism and/or “copy-paste” has reduced to zero

Figure 2. Students performance—Re-submission

Discussion

It has become a routine practice amongst several UK universities to automatically check all (or important) student submissions for plagiarism. In fact Turnitin® reports are being used as absolute evidence in academic disciplinary hearings. Also there is evidence that indicates the use of anti-plagiarism software led to a decrease in plagiarism (Batane, 2010). However less data is available on the effectiveness of this software as a teaching tool (Bilic-Zulle et al, 2007). Therefore the main objective of this study was to assess the potential of Turnitin® software as a self-teaching tool. The study was carried out as a part of on-going anti-plagiarism activity workshops aimed at level 1 undergraduates.

It is reassuring to note from the data that most students have a positive attitude towards academic writing and avoiding plagiarism. Approximately 53% of (37 out of 70) students were able to write this scientific report with proper attributions and referencing. Secondly, by highlighting the matches with original sources, the software has clearly helped the students to learn their mistakes. Interestingly, the data from second submissions have shown that at least 59% of (22 out of 37) students were able to re-write their reports without plagiarising. The rest (15 out of 37 students) showed improvements in their academic writing. Thus allowing them to view their submission reports helped them to identify their mistakes. The exercise also helped the author to identify the “needy” students who required extra help in avoiding plagiarism.

Manual proof reading and cross matching their first and the second submissions has highlighted several positive as well as negative aspects of using the software as an educational tool. On the positive side, it has highlighted the common mistakes made by the students while paraphrasing. Especially, students were able to identify the mistakes in using and/or attributing ideas from published sources. As the result of this they were

Table 1
Common student misconceptions and their justifications

Common Misconceptions	Justifications
Patch-work is not plagiarism.	It has been “modified” and properly referenced.
Figures from the internet need not to be referenced.	They are there to be freely used.
Information from abstracts can be copied	That is not the actual article
It is alright to collude with the student with whom “paired/group” was carried out.	This is what the “group/paired” work means
There is no need to reference non-copyrighted articles	Care should only be taken when “copying” copy-righted articles
Re-using essays or “text recycling” (self-plagiarism) is acceptable	It was me who wrote it in the first place
Information from the secondary source can be used without checking their validity	It is hard to get hold of the primary source
Paraphrasing means using sentences from the original article in inverted commas.	There are only a few ways the information can be re-written.

able to either address the mistake or seek academic help (Wicker, 2007). In other words it has enhanced reflective learning. This confirms the findings from a questionnaire-based study conducted by Whittle and Murdoch-Eaton in 2008. In their study students claimed that it was “reassuring to check” for plagiarism before submission. The main difference in this study from Whittle and Murdoch-Eaton’s results is the students were given a chance to improve their reports by allowing them to re-submit. Interestingly, this re-submission opportunity has been found to be a controversial subject amongst the academics, some supporting it with caution and recommending that it should be carefully monitored and coupled with proper guidance whilst others rejected it outright.

The later group’s argument was although this practice can be regarded as “feedback” to improve academic writing, this might also allow them to learn or discover the places where the software failed to highlight sentences which they knew they had plagiarised. This would eventually lead to students learning about faults in the system so potentially they could try to evade detection using these “holes” and ultimately make the software less useful as a teaching tool. This is a valid point as Turnitin[®] (like any other automated system) has limitations. On the accuracy of detecting the source, the software correctly matched most articles that are available on the internet. However the matching rate was low when students used information from the PubMed[®] abstracts. This may be due to the fact that the software cannot access this web page, as explained in Turnitin[®] software itself (www1). However this is worrying as PubMed is one of

Table 2

Examples excerpts from students' first and re-submissions

Excerpt from First submission	Potential academic irregularity	Modified Re-submission
<p>"<i>E. Faecalis</i> is a known pathogen and is known to colonise the large intestine and the urinary tract (Rollin, 2000). This is helped by the fact that its optimum temperature is close to the optimum of humans (Brooker et al, 2008)"</p>	<p>"Patch-work" using information from two books, without forming the link.</p>	<p><i>E. Faecalis</i> is a pathogen that can cause intestinal and urinary tract infections (Rollin, 2000). Since the optimum temperature needed for its growth is approximately the same as human body temperature, it can easily colonise human intestine (Brooker, 2008).</p>
	<p>*Academic advice was given how to make a meaningful sentence, in context with the assignment title</p>	
<p>...laboratory blood assays designed to give information about the state of a patient's liver, called liver function tests. The parameters measured include albumin, bilirubin and others. Liver transaminases (AST/ALT) are believed not to be liver function tests, but biomarkers of liver injury in a patient with some degree of intact liver function.</p>	<p>Cut & paste plagiarism. Information lifted, word-for-word from www.medlibrary.org</p>	<p>According to www.medlibrary.org "<i>liver function tests are laboratory blood assays designed to give information about the state of a patient's liver</i>". "<i>The parameters measured include albumin, bilirubin and others. Liver transaminases (AST/ALT) are not liver function tests, but biomarkers of liver injury</i>" (see reference)</p>
	<p>*Academic advice given but student was not fully engaged with the process. This can be seen in the re-submission which shows student tried to acknowledge the source but has not achieved the learning outcome.</p>	
<p>Immunocytochemical analysis by TH showed a significant decrease in TH-positive cells after 1 mM MPP+ for 24 h, and the surviving TH-positive had an immature appearance with an apparent loss of processes. Scale bar: 400µm for 5X and 100µm for 20X. Inset: An example of TH-positive cells in a colony stained with DAPI to show cell nuclei (Freed, 2006)</p>	<p>Inappropriate paraphrasing when student tried to narrates data (and a figure) from published article.</p>	<p>Using immunocytochemical analysis Freed (2006) has shown that after 24 hours treatment of 1 mM of MPP+ has significantly reduced the tyrosine hydroxylase (TH) positive cells. Interestingly the surviving TH-positive had an immature appearance. Figure adapted from Freed (2006):- Scale bar: 400µm for 5X and 100µm for 20X. Inset: An example of TH-positive cells in a colony stained with DAPI to show cell nuclei</p>
	<p>*Student did not seek academic advice but tried to modify the statement with proper paraphrasing and attributions. Her re-submission shows the student has tried to understand the article and then paraphrased.</p>	

One example each from (1) patchwork (2) cut & paste plagiarism and (3) inappropriate paraphrasing from the first submission is given in the left hand column. Their improvements in re-submissions are given in the respective right hand column. Additional information is given in the middle column.

the commonest open access websites used by the scientific community, from students to researchers. In the case of sentences taken from published scientific books or even recently published articles, the performance by the software is even lower. Turnitin[®] failed to identify any of the material plagiarised from the books. Consequently allowing students to access the report may result in them finding these “holes”. However, this can be addressed by limiting the opportunity of students, to check the Turnitin[®] reports to one attempt. Moreover, it should be noted that the final submission is still going to be checked by the academics. As long as the academic “wisdom” and manual checking is applied, any attempts to cheat the system can be minimised. In fact this would give a strong message to the students that their reports are carefully scrutinised not only electronically but also manually checked—a better deterrent against plagiarism.

Considering the student misconceptions, it is clear that they are confused about the differences between (a) patchwork; (b) paraphrasing; and (c) quoting. Although these were addressed with relevant examples in the anti-plagiarism workshops, perhaps they did not grasp the subtle differences and need to learn by making these mistakes. Opportunities to check their submissions would definitely enhance their understanding.

By matching the sentences, Turnitin is actually showing the students that they have copied and therefore makes them think whether they have achieved the learning outcomes.

At least it would help the students understand that a properly acknowledged (and/or quoted paragraph) may not be considered as plagiarism; but has still not fulfilled the learning outcomes (and would therefore attract low marks). Using these examples, academics can effectively enhance their students’ reflective learning and also make them learn to attribute ideas in context with the research question.

Conclusions

The data from this study suggest that Turnitin[®] can be used as a teaching tool to enhance reflective learning, provided the process is controlled properly by the academics. It is also clear that the exercise should be linked with anti-plagiarism workshops and/or personal tutoring systems to assist the students’ understanding and therefore helping them learn from their mistakes.

References

- Australian Universities Teaching committee (2007) Plagiarism detection software: How effective is it? In *Assessing Learning in Australian Universities: ideas, strategies and resources for quality in student assessment*. (www.cshe.unimelb.edu.au/assessinglearning, Accessed on 20th December 2012)
- Badge, J. & Scott, J. (2009) Dealing with plagiarism in the digital age. *School of Biological Sciences, University of Leicester*. (<http://www.heacademy.ac.uk/assets/evidencenet/syntheses/leicester.pdf>, Accessed on 04th April 2013)
- Batane, T. (2010) Turning to Turnitin to Fight Plagiarism among University Students. *Educational Technology & Society*, 13 (2): 1–12
- Beckton, J. (2007) Student access to Turnitin, Plagiarism discussion list, 1 February [Online]. Available in (Educational resource review) DOI:10.3794/johlste.82.res (www.heacademy.ac.uk/johlste)

- Bilic-Zulle, L., Azman, J., Frkovic, V. & Petroveck, M. (2007) Is There an Effective Approach to Deterring Students from Plagiarizing? *Sci Eng Ethics* 1471–5546 (Online version).
- Cheah, SW. & Bretag, T. (2008). Making technology work for academic integrity in Malaysia. Paper presented at the *3rd International Plagiarism Conference*. Northumbria University, UK, 21–23 June 2012
- Davis, M., & Carroll, J. (2009). Formative feedback within plagiarism education: Is there a role for text-matching software? *International Journal for Educational Integrity*, 5(2), 58–70. (Retrieved from <http://www.ojs.unisa.edu.au/index.php/IJEI/article/view/614>)
- Dhal (2007) Turnitin®: The student perspective on using plagiarism detection software *Active Learning in Higher Education* 2007; 8; 173. (Retrieved from <http://alh.sagepub.com/cgi/content/abstract/8/2/173> Accessed on 2nd April 2013)
- Guader, H. (2004) The Hunt for Plagiarism, *Internet Reference Services Quarterly* 8(4): 27–41
- Jocoy, C. & DiBiase, D. (2006) Plagiarism by Adult Learners Online: A case study in detection and remediation. *The International Review of Research in Open and Distance Learning*, 7(1). (Retrieved from <http://www.irrod1.org/index.php/irrod1/article/view/242/495>)
- Jones, M. (2008) Essays for sale: time for legal regulation? *Third International Plagiarism Conference*, Newcastle upon Tyne, UK, 23–25 June 2008
- Nicol, D., & Draper, S. (2008) Redesigning written feedback to students when class sizes are large (pdf). Paper presented at *the Improving University Teachers Conference, 29 July–1 August, Glasgow*
- O'Hara, M., Caron, C., & Gillian, M. (2007) Getting it write: a multidisciplinary approach to student academic support. Available on-line: http://www.plagiarismadvice.org/documents/ohara_casestudy.pdf (accessed on 4th April 2013).
- Rowell, Carroll, Morris & Jameson (2009) Educational resource review. *Journal of Hospitality, Leisure, Sport and Tourism Education*, 8(2):157–166
- Satterwhite, R. & Gerein, M. (2001) Downloading detectives: searching for on-line Plagiarism. (Available at <http://www.coloradocollege.edu/library/course/webplag2.html>, accessed on 3rd April 2013).
- Sivasubramaniam, SD. (2004) Plagiarism amongst biomedical science undergraduates—concerned academics versus ignorant students. *Proceedings on Plagiarism: prevention, Practice & Policy* (eds) Smith AP and Duggan F, 205–224.
- Sivasubramaniam, SD. (2006) Assisting students to avoid plagiarism: The role of formative workshops. *2nd International Plagiarism Conference (Plagiarism Advisory Service)* 32.
- Wicker, P. (2007) Plagiarism: understanding and management. *J Perioper Pract.* 17(8):372, 377–82.
- Whittle, SR. & Murdoch-Eaton, DG. (2008) Learning about plagiarism using Turnitin detection software. *Medical Education* 42: 513–543.
- Wright, D., Owens, A. & Nigel, D. (2008) Making the case for multiple submissions to Turnitin. paper presented at the *3rd International Plagiarism Conference*. Northumbria University, UK, 21–23 June, 2012. WWW1: <http://www.jiscpas.ac.uk/faqsearch.php?sol=turnitin> (Accessed on 20th December 2012)

Author

Dr S. D. Sivasubramaniam, shiva.sivasubramaniam@ntu.ac.uk, School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK

Copyright © 2013 Authors listed on page 149 grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this

document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

INVESTIGATING ACADEMIC PLAGIARISM: A FORENSIC LINGUISTICS APPROACH TO PLAGIARISM DETECTION

Rui Sousa-Silva

Abstract: Automatic plagiarism detection tools have evolved considerably in recent years. Owing in part to the recent technological developments, which provided more powerful processing capacities, as well as to the research interest that plagiarism detection attracted among computational linguists, results are nowadays more accurate and reliable. However, most of the plagiarism detection systems freely and commercially available are still based on similarity measures, whose algorithms search for identical or, at most, similar strings of text, within a more or less short search distance. Although these methods tend to perform well in detecting literal, verbatim plagiarism, their performance drops when other strategies are used, such as word substitution or reordering. This paper presents the results of a forensic linguistic analysis of real plagiarism cases among higher education students. Comparing the suspect plagiarised strings against the most likely originals from a legal perspective, it is demonstrated that strategies other than literal borrowing are increasingly used to plagiarise. A forensic linguistic explanation of the strategies used, and why they represent instances of plagiarism, is then offered. The paper concludes by indicating some improvements to increase the accuracy and reliability of the detection procedure.

Plagiarism and Forensic Linguistics

Plagiarism, which in its most basic form consists of passing off someone else's work as one's own, has attracted considerable media attention in recent years, mostly due to the high profile of people involved. Examples include the case of the German Defence Minister Karl-Theodor zu Guttenberg, who, in 2011, (temporarily) renounced his doctorate title and eventually resigned, as a result of accusations that he had plagiarised when writing his doctoral thesis. In Romania, the Prime Minister Victor Ponta was accused, in 2012, of plagiarising substantial portions of his doctoral thesis, and faced pressure to resign. More recently, this year, suspicion was raised by plagiarist hunters that the German Education Minister Annette Schavan had plagiarised at least 130 passages of her doctoral thesis; as a result of the suspicion, the University of Düsseldorf, which had awarded her PhD, revoked her title, after conducting an official process to rescind it, and she later resigned. A few years earlier, a journalist of the Portuguese quality newspaper *Público* was accused of plagiarising Wikipedia and the *New Scientist*, and more recently the journalist of *The Independent* Johann Hari was suspended for plagiarising news articles. Cases of academic plagiarism are also known. In 2010, for example, a Portuguese university lecturer resigned following accusations that she had plagiarised her doctoral thesis. However, in the academy not all cases make it to the news. Rather on the contrary, most cases tend to be resolved internally, by disciplinary boards or the lecturers/tutors themselves, depending on the respective institution. Academic plagiarism is, nonetheless, considered an unacceptable practice,

which brings along severe penalties, up to having their titles rescinded, even if the instances of plagiarism are not found until a later date.

In these, as in most cases, plagiarism is seen, not only as an immoral, but also as an illegal action, whose nature legitimates punitive actions that include, among others, rescinding titles. But these considerations of plagiarism both as an immoral act and an illegal action bring additional problems, the most challenging of which consists of determining the degree of intentionality underlying the instance of plagiarism. As Howard (1995) claimed, Angèlil-Carter (2000) later argued and Pecorari (2008) subsequently reiterated, academic plagiarism is more often a problem of academic writing skills (or their lack thereof), than an intentional attempt at passing off someone else's work as one's own. Likewise, Scollon (1994; 1995) and Thompson (2002) sustained that non-compliance with academic writing procedures and conventions was often more a result of clashing intercultural aspects, than an intention to deceive. Additionally, if text reuse is taken to represent a form of authorship, as Robillard (2008) argues, then a clear distinction has to be made between improper, unintentional borrowing, and intentional plagiarism (Howard 1995). In their research, both Angèlil-Carter and Pecorari attempted to determine the plagiarists' intention by identifying the instances of textual borrowing and then interviewing the suspect plagiarists. They concluded that, in some cases, the students misattributed their sources inadvertently, whereas other cases suggested that the plagiarists acted with the intention to deceive.

Determining the suspect's intention by interviewing them, however, may not always be a possible investigative method. Firstly, due to the reported increasing number of plagiarism cases, most universities will rarely have sufficient human and technical resources to investigate all cases thoroughly and properly. Secondly, if we consider that some instances of deception pass unnoticed even in courts of law, depending on the 'expertise' of the deceivers, then lecturers/tutors and educational institutions can hardly be expected to properly identify all instances of deceptive plagiarism. It adds to this that the plagiarist may himself/herself misjudge their case, either by wrongly admitting the truthfulness of false positives, or by denying the truthfulness of true positives. Not to mention the need for proper evidence that proves the claims for—or against—plagiarism. Finding evidence raises specific challenges, whether it is a case of plagiarism (where a text borrows from (an) other source(s) without acknowledgement) or collusion (where two or more people work collaboratively on the same text and pass off each individual document as an original), especially when the plagiarist has practiced a deceptive act whose nature results from lying (Eiras & Fortes 2010). Firstly, as Eggington (2008) concluded, deception can hardly be detected linguistically; secondly, as Coulthard & Johnson (2007) argued, it is not the linguist's task to detect the plagiarist's intention; on the contrary, they sustained that it is the linguist's task to establish whether two texts have been produced independently or otherwise. Analyses of this type, which are based on the comparison of suspect texts and potential originals, have been used in academic, as well as nonacademic contexts, and are at the basis of most plagiarism detection software packages. However, as a result of the technological developments of the last decades—especially the Internet—more information is now more readily available, including to students, which makes it easier to pass off someone else's work as one's own, by copying and pasting the original text 'as is', or by making

minor or more substantial alterations to it. At the same time, due to the massive volume of information available, it is now more difficult for any reader to intuitively identify a text or text passage as an instance of plagiarism. But as Coulthard & Johnson (2007) argue, the technological developments that made it easier to plagiarise, also made it easier to detect instances of plagiarism. The need to detect instances of plagiarism that are missed by intuition, together with these recent technological developments, particularly in computer sciences and computational linguistics, and the growing interest of computer scientists and computational linguists alike, led to the development of a plethora of plagiarism detection software packages.

Existing plagiarism detection software can operate based on two different approaches: *external plagiarism detection* and *intrinsic plagiarism detection* (Potthast et al. 2009). The latter aims to detect instances of plagiarism in cases where the reader is intuitively led to the suspicion that the text has been borrowed from other sources, but does not know any original texts against which the text can be compared. The detection procedure is, in this case, based on an intrinsic, stylistic analysis of the suspect text, in order to identify stylistic inconsistencies that can be used to challenge the authorship. Although this procedure may represent a valuable contribution, from an investigative perspective, by not presenting the original source from which the text was lifted, it lacks the evidential value required to demonstrate the instance of plagiarism. Most common plagiarism detection software packages currently available therefore operate via an external analysis, by establishing a comparison between the suspect text(s) and the known originals, in order to determine the degree of similarity or identity between the texts. This procedure, which is used (even if with minor or major adjustments) by most detection systems—including Turnitin and SafeAssign—works by scanning the texts and applying computational string-matching techniques to identify words, phrases, sentences or paragraphs that, having been copied and pasted from another source ‘as is’, or subsequently altered, are identical or similar to the original text. Systems that use this approach perform well in detecting identical texts, based on *verbatim*, word-for-word borrowing, but not when changes are introduced to the original text. In this case, the detection gradually becomes more difficult to handle computationally, up to a point where it becomes impossible.

This problem imposed on the computational detection of plagiarism is due mainly to search space restrictions. Since any two texts are expected to share a high number of words, most of which are grammatical (and consequently used less ‘uniquely’), flagging all individual items that are shared between two texts will lead to the wrong identification of an instance as plagiarism. Therefore, search space restrictions have been introduced to ‘teach’ the system that not all overlapping words should be flagged as plagiarism; on the contrary, the system is instructed to flag as plagiarism only overlapping strings of co-occurring words of a certain length in the original and suspect texts. By determining the minimum number of words that must co-occur, as well as the maximum number of new words that are altered, introduced or deleted from the string, before a text can be considered an instance of plagiarism, the system avoids flagging *false positives*, i.e. where strings of plagiarism are misidentified. Consequently, if a string of overlapping text is below a certain number of words, or if the number of words that are altered, deleted from or introduced to the original text is above a

certain threshold, the system traditionally identifies it as original text. This raises some problems. Firstly, as Woolls (2010) explains, and as is commonly advertised by plagiarism detection software packages, the volume of overlapping text that is calculated usually requires a manual, human analysis, in order to confirm or otherwise reject a certain flagged instance as plagiarism. On the other hand, if we consider that, the more an unattributed text is manipulated, the higher the plagiarist's intention to plagiarise (Sousa-Silva et al. 2010), then the more a text is altered, the more severe the instance of plagiarism, and the lower the likelihood that it will be identified by the machine.

Linguistically grounded approaches are therefore required, not only to raise suspicion, but also to provide evidence that a text has been lifted from another source, to explain the linguistic strategies adopted, and additionally to assist lecturers/tutors and disciplinary board members, among others, in determining the plagiarist's intention. The type of linguistic analysis conducted by forensic linguists has demonstrated good results in this respect. Although it is often considered that the impact of academic plagiarism is limited to the academy, the cases discussed above demonstrate otherwise, i.e. the impact of academic plagiarism can and does reflect in non-academic contexts, to challenge the qualifications, or, not the least, the ethics of highly renowned professionals whose titles are rescinded as a result of previous actions. In these cases, suspicion often suffices to socially impact the suspect plagiarist's life, but solid evidence is required to legally support the decision adopted, especially when this involves definite actions up to rescinding or revoking a title. Research into forensic linguistics, which consists of applying linguistic methods and analyses in forensic contexts, has been used effectively in cases of fraud where linguistic evidence is vital, and has demonstrated that the likelihood that a text—or set of texts—has not been produced independently can be determined accurately. Moreover, as has been demonstrated (Turell 2008), such data can be used, not only as an investigative tool, but also as evidence.

The purpose of this study is to challenge the assumption that plagiarism detection software can effectively identify the most serious instances of plagiarism, where the plagiarist has heavily and intentionally manipulated the text to deceive his/her readers. Using a combination of descriptive linguistic analyses of instances of academic plagiarism, this study presents some cases that, owing to their nature, could be missed, in whole or in part, by plagiarism detection systems. This study indicates that *word substitution* and *reordering*, as well as *translation*, are some of the strategies used by plagiarists to mislead the detection systems.

This paper is structured as follows. The following section explains how the research is operationalized; it describes the corpus of texts analysed in this study and the analytical method employed. The findings of this analysis are presented in the subsequent section, which is followed by a discussion of the findings. The paper concludes with a summary of the findings, and with a discussion of future research directions.

Table 1
Assignments included in the corpus

Student	Number of Words
S1	3,638
S2	1,370
S3	3,333
S4	4,629
S5	2,033

Method of Analysis

The analyses of instances of plagiarism commonly consist of comparing suspect texts against the putative originals, and highlighting the textual identities and similarities, or, alternatively, the differences between the texts. One can hypothesise that, the higher the identity between the derivative and the original texts, the easier it is to detect the instance of plagiarism, and the easier the machine detection. Conversely, the higher the number of edits introduced to the derivative text, the more difficult the detection procedure becomes, especially when using detection software. To test this hypothesis, a small corpus of academic assignments that were found to have plagiarized was used to conduct an extrinsic analysis. The assignments were written in Portuguese by post-graduate students of two Portuguese universities, on media studies (S3, S4 and S5), and on design (S1 and S2). A corpus of texts written in Portuguese offers an additional advantage, when compared to English: since Portuguese is morphologically and syntactically more diverse and flexible than English, it offers a greater range of word combinations and inflections, and consequently raises new challenges to the detection procedure. The original sources were also provided by the lecturers/tutors, for comparison.

As shown in Table 1, these assignments are of considerable length (an average of 3,000 words per essay).

However, for the purposes of the analysis of the linguistic features used to plagiarise, or to assess the impact of these strategies on the detection procedure, this quantification is irrelevant. This is because, on the one hand, there is no correlation between the number of words and the amount of borrowing, and, on the other hand, between the text size and the linguistic strategies used.

The linguistic analysis focused on the nature of the instances that showed changes, in terms of *word substitution*, *word reordering* and *translation*. Since the aim of this study was to identify the nature of the changes operated, no detection software was used at this stage. The potential impact of these alterations on the manual and software detection is explored in the descriptive analysis of the data. A manual, side-by-side comparison between the original and the derivative text was made, highlighting alterations in grammar, punctuation, syntax, semantics, lexis and discourse. Since the derivative texts were, for the most part, borrowed *verbatim* from the original, the differences, rather than the similarities, were highlighted to signal the alterations

introduced, and the identical strings, showing exact matches, were discarded. The next step consisted of the descriptive linguistic analysis of the strings that had been altered by replacing or reordering the words of the original. Subsequently, the strings that had been translated from the original source were analyzed more closely. Finally, those alterations, and specifically their relevance to determining the impact on the machine detection procedure, were investigated, in order to determine whether they are to be expected or, on the contrary, whether they are illicit.

Results of the Analysis

The first stage of the analysis consisted of identifying the strings of text containing *word substitution*, *word reordering* and *translation*. Although some of these linguistic strategies are often used to paraphrase, reference to paraphrasing is avoided in this study. This is because paraphrasing involves a deeper rephrasing that goes beyond the three types of alterations discussed, in order to retain the meaning, while using a new form.

Word Substitution

Word substitution consists of replacing a word or combination of words with words with identical or similar meaning. Although these replacement words usually retain some sort of semantic relationship with the original text (such as synonymy, hyponymy or hypernymy), they can also be from a different semantic field, especially when they aim to retain the coherence with the extra-textual world. The assignment of S1 presents several instances of the latter. The word ‘escola’ (*school*) in the original is replaced with ‘cultura’ (*culture*) in the derivative text; ‘um cantor ou uma atriz’ (*a singer or an actress*) is replaced with ‘um designer ou um artista plástico’ (*a designer or a plastic artist*); ‘os professores e os pais’ (*the teachers and the parents*) is replaced with ‘os profissionais e o publico em geral’ (*the professionals and the general public*); ‘educativa’ (*educational*) is replaced with ‘cultural’ (*cultural*); ‘um jogo de futebol’ (*a football match*) is replaced with ‘a performance de um artista’ (*an artist’s performance*), and ‘jogo’ (*match*) and ‘partida de futebol’ (*football match*) are replaced with ‘performance’ in both instances. Word substitution is not, however, used exclusively in this assignment. S5, for example, replaces the word ‘mesclado’ (*entangled*) with the synonym ‘embaralhado’, ‘fotodocumentalismo’ (*photo documentary*) with the semantic equivalent ‘fotodocumentário’, and ‘exigem’ (*demand*) with the semantically related ‘necessitam’ (*require*). In this same assignment, the adjective ‘sustentada por’ (*argued by*), followed by the author’s name, is replaced with the semantic equivalent preposition ‘To’, followed by the author’s name. S4’s assignment also shows many cases of word substitution. But, interestingly, most of these are minor, i.e. they result either from the correction of Brazilian Portuguese spelling to the European Portuguese variant, or reflect the different use of prepositions in the two variants. However, there are also substantial lexical substitutions, whose nature involves more than simple spelling checks. For example, ‘enxergar’ (*see*) is replaced with the synonym ‘olhar’, ‘superposição’ (*overlapping*) is replaced with its equivalent ‘sobreposição’, and ‘plasmar’ (*exhibit*) is replaced with the synonym ‘passar’.

A more substantial change is operated by the substitution of ‘vermelhos’ (*reds*) with ‘cores vermelhas’ (*red colours*) in the derivative text.

A sophisticated substitution is operated by S₁ in the phrase ‘opção que condicionará’ (*an option that will condition*). Originally used as a subordinate clause, this phrase is reused in the derivative text as part of a new sentence: ‘Esta opção irá condicionar’. A new demonstrative is introduced (‘Esta’), which had been omitted in the original, the relative pronoun ‘que’ (*that*) is deleted from the derivative, and the future tense of the verb to condition, ‘condicionará’, is replaced with the infinitive form of the verb, ‘condicionar’, preceded by the future tense of the auxiliary verb ‘ir’. These alterations result in a new wording that, although semantically identical to the original, is morpho-syntactically different, and sufficient to trick machine detection systems.

In some cases, words are used to replace punctuation. S₄, for example, replaces the semi-colon with the adversative ‘mas antes’ (*on the contrary*). Likewise, in S₅’s assignment, ‘Mais:’ (*additionally*), whose specific meaning in this context is marked by the use of the colon, is replaced with the lexical equivalent ‘Depois’.

Word Reordering

Word reordering is used to describe the linguistic operations whereby the original words are reused, but in a different. Although this linguistic strategy is not as common as word substitution, the corpus includes several examples. S₄ uses this linguistic device as a plagiarism strategy several times: ‘se deixar envolver’ (*let themselves involve*) is replaced with the more European Portuguese standard ‘deixar envolver-se’; ‘que foram assim chamados por’ (*that were thus called by*) is replaced with ‘assim foram chamados pelos’; the Brazilian Portuguese syntax ‘ele se separaria’ (*he would depart*) is replaced with the European Portuguese ‘separar-se-ia’. S₅ also uses this strategy: ‘No sentido lato, entendemos por fotojornalismo’ (*In the general sense, by photojournalism we mean*) is reordered ‘por fotojornalismo no sentido lato, entendemos’. Even more complex is the reordering operated in Example 1 below, where the original:

Adaptando ao fotojornalismo uma sistematização das funções da linguagem no discurso informativo sustentada por Jesús González Requena(41), poderíamos (...)

is reordered:

Para Jesús Requena, adaptando ao fotojornalismo uma sistematização das funções da linguagem no discurso informativo poderíamos (...)

Example 1: A sophisticated case of reordering

In this case, the name of the author is edited (*González* is deleted), the comma (as well as the note number) is deleted, and the reporting phrase (‘sustentada por Jesús González Requena’) is altered (‘Para Jesús Requena’). As a consequence, a maximum of 11 running words are retained in the derivative text, five of which are grammatical items, of a chain of 25 running words.

Translation

Finally, *translation* is where a writer has an original translated from another language, via human or machine translation, and uses it as his/her own original, while omitting the source. An extensive example of this strategy is provided by S3's assignment. This assignment includes a literal translation of an original in Spanish into Portuguese that retains, for the most part, the original punctuation, lexis and syntax. Besides some spelling mistakes ('aerosois' instead of 'aerosóis' for *aerosols*), this assignment shows several other mismatches. In terms of lexis, some words are wrongly used, either because they do not exist in Portuguese (e.g. 'decoraciones', 'carácter espontáneo'), or because, when existing, they have a different meaning (e.g. 'pintada', 'mural', 'rótulos'). Moreover, as a simple search using a common Internet search engine demonstrates, the phrase 'escritor de graffiti' is common in Spanish, but not in European Portuguese. Likewise, some phrases like 'Pois bem', 'Assim mesmo', 'Agora bem', and 'à hora' as a translation of 'Pues bien', 'Así mismo', 'Ahora bien', and 'a la hora', respectively, indicate a wrong literal translation. Syntactically, a wrong transfer is noted in phrases like 'ia alguém a se arriscar' as a translation of 'iba alguien a arriesgarse'. In terms of grammar, this assignment consistently shows a wrong use of uppercase after a colon; although this may be common in Spanish, in Portuguese lowercase is to be expected after a colon. Additionally, there are some problems with the concordances; for example, the phrase 'Numerosos foram as tentativas' is grammatically wrong in Portuguese because, as this is a gender-sensitive language, 'Numerosas' (instead of 'Numerosos') is expected, in order to retain the consistency with 'tentativas'. In Spanish, however, since 'intentos' is a masculine noun, 'Numerosos' is used.

Discussion

An investigation into plagiarism needs to consider the particular circumstances involved. If, as Howard (1995), Angèlil-Carter (2000) and Pecorari (2008) argued, plagiarism among students is a pedagogical problem that can be, for the most part, resolved by teaching the students how to write academically, then we have to agree that academic and non-academic plagiarism cannot be judged independently of their circumstances. Specifically, in the academy, where instances of plagiarism represent a failed attempt by students at writing academically (Howard 1995; Pecorari 2008), plagiarism can result from a legitimate attempt at producing a good piece of writing, or, alternatively, an attempt at obtaining the best possible grade, for the minimum effort. Consequently, if one considers that the principle behind plagiarism is laziness, then only minor alterations are to be expected, as these do not require hard work. This is the case, for example, where only one word or a few words are altered in a long sentence.

Minor alterations of this type do not impact the machine detection procedure, since they are not sufficient to interrupt the minimum chain required from the string matching procedure. Conversely, the machine detection procedure is made more difficult in cases where the alterations introduced break the chain of consecutive words in such a way that the sequence of running, overlapping words is not sufficient to run the detection procedure against an unknown source. Example 1 above illustrates

this point well. Taken together, the alterations introduced to the original sentence transform a total of 25 running words into a chain that is broken down into three batches of overlapping words: 11, 1 and 4 words, respectively. Moreover, considering the principle of lexical richness, as discussed by Coulthard & Johnson (2007), even the longest string of these (11 words) loses significance owing to the fact that roughly half of the words are grammatical items, and hence more likely to occur anyway.

Punctuation is another element that impacts the machine detection procedure. In order to avoid the maximum number of false positives, while at the same time attempting at identifying true cases of plagiarism, some detection software packages use strong punctuation to divide the text into chunks. Consequently, the detection procedure is affected in cases where words are used to replace punctuation, as is the example where the semi-colon is replaced with the adversative 'mas antes'. Machine detection is also hampered in cases where, after dissecting the original into smaller sentences, the plagiarist substitutes at least some of the original words, as in the example 'opção que condicionará'. In this case, the amount of consecutive words that are shared with the original text is so small, that the software can hardly identify the string as plagiarism. Likewise, the detection procedure is also impacted by the addition of new words, together with word substitution or reordering. The phrase 'que foram assim chamados por', discussed above, illustrates this point well. The derivative sentence is not only split by the reordering, but also 'Movimentos como' (Movements such as) is added to the beginning of the sentence, and a sequence of 8 running words is interrupted by the determiner 'o' (the), resulting in 5 and 3 running words, respectively; finally, a sequence of 9 words is added to the end of the sentence. Taken together, these alterations impact the machine detection procedure, not only by interrupting the chain of consecutive words, but also by increasing the ratio of new words, in relation to the words of the original. As a consequence, the number of reused words, in this particular case, may be lower than the threshold required by the detection system to flag a text as plagiarism, and therefore falsely considered to be original text.

Translation also represents additional problems to plagiarism detection, starting with the definition of plagiarism. Specifically, translation can be considered a plagiarism strategy if plagiarism is defined as passing off someone else's works and ideas as one's own, but not if the restrictions imposed on borrowing apply only to words. Since a translation involves a transfer of the meaning of an original in one language using the linguistic signs of another language, these signs are necessarily different from the original ones. Consequently, the text that is lifted from another original is not similar, and much less identical. This represents a problem to computer systems, which need to process texts using comparable patterns to be able to proceed to the string matching. In this case, it is a requirement that the two texts are converted to one common language for comparison. However, this conversion is only possible if the original is known, which requires that the reader either (a) knows the original text, or (b) the text provides linguistic cues that lead the reader into intuitively establish the language of the original. These cues are usually provided by issues in grammar, punctuation, syntax or lexis, such as the ones discussed in the examples illustrated above. However these cues can be discounted when the writers are known to be writing in a foreign language, in which case issues with grammar, punctuation, syntax or lexis

are to be expected. The challenges imposed by translation on the detection procedure are even bigger when this strategy is used in combination with other strategies, such as word substitution or reordering.

The combination of strategies is, speculatively, one of the major challenges imposed on software detection systems. Different plagiarism detection software packages have demonstrated different degrees of effectiveness in detecting different plagiarism strategies. Some packages (e.g. Turnitin) perform well in detecting identical text, regardless of the nature of the words (lexical or grammatical), whereas others (e.g. SafeAssign) offer the users the possibility of excluding certain strings from the plagiarism report, and others (e.g. CopyCatch) focus on the lexical items to calculate the percentage of plagiarised lexical vocabulary. However, as the analysis of these corpus texts demonstrates, it is very unlikely that only one strategy is used individually when plagiarising; on the contrary, a combination of plagiarism strategies within the same text is not uncommon. Since, at this stage, it is computationally challenging, if not unrealistic, to combine the possibility of detecting several plagiarism strategies within one same detection system, software packages have until now given priority to one or other strategy. Ascertaining, ‘beyond reasonable doubt’, that the suspect text is a derivative of the original therefore requires the manual analysis of a human ‘detector’ (ideally a trained forensic linguist), who is able, as Woolls (2010) argues, to handle the complexity underlying the principle of similarity.

Conclusion

This paper discussed specifically three linguistic strategies used to plagiarise: word substitution, word reordering and translation. It demonstrated, with examples from a small corpus of instances of plagiarism, that these strategies are commonly used to plagiarise, and that at least some amount of editing is expected from instances of plagiarism, not the least as a result of proofreading.

A linguistic analysis of these instances, identical to that applied in forensic contexts, was provided, on the one hand to describe how these strategies were operationalized, and, on the other hand, to explain why they represent plagiarism. This analysis indicated that, if one considers the limitations of plagiarism detection software, some of these instances of plagiarism can be misidentified as original text, especially when the text is altered substantially, or when a combination of strategies is used. The latter, especially, has the potential to hamper the machine detection and pass unnoticed, even if it is potentially the most relevant in demonstrating the plagiarist’s intention to consciously manipulate the text and pass it off as his/her own.

Some improvements to software detection were suggested that could increase the effectiveness of the machine detection procedure, including the ability to handle a combination of different strategies, or the need to be agnostic to the word order when building an index of the words used. However, some of these improvements may take some time to be implemented, especially owing to the requirements in terms of processing capacities. Additionally, as Woolls (2010, p.590) argues, ‘any computer program can only be an approximation of what human readers can recognise and handle’. It is therefore relevant that, in order to avoid the misclassification of

instances of plagiarism, the analyses and the reports provided by detection systems are interpreted with the assistance of a forensic linguistic analysis, so as to discard, on the one hand, false positives, while at the same time unveiling hidden true positives that may have been missed by the detection systems.

References

- Angélil-Carter, S., 2000. *Stolen language?: plagiarism in writing*, Harlow: Longman.
- Coulthard, M. & Johnson, A., 2007. *An Introduction to Forensic Linguistics: Language in Evidence*, London and New York: Routledge.
- Eggington, W.G., 2008. Deception and fraud. In J. Gibbons & M. T. Turell, eds. *Dimensions of Forensic Linguistics*. Amsterdam and Philadelphia: John Benjamins, pp. 249–264.
- Eiras, H. & Fortes, G., 2010. *Dicionário de Direito Penal e Processo Penal*, Lisboa: Quid Juris.
- Howard, R., 1995. Plagiarisms, Authorships, and the Academic Death Penalty. *College English*, 57(7), pp. 788–806.
- Pecorari, D., 2008. *Academic Writing and Plagiarism: A Linguistic Analysis*, London: Continuum.
- Potthast, M. et al., 2009. Overview of the 1st International Competition on Plagiarism Detection. In B. Stein et al., eds. *SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09)*. CEUR-WS.org, pp. 1–9.
- Robillard, A.E., 2008. Situating Plagiarism as a Form of Authorship: The Politics of Writing in a First- Year Writing Course. In R. Howard & A. Robillard, eds. *Pluralizing Plagiarism: Identities, Contexts, Pedagogies*. Portsmouth: Boynton/Cook, pp. 27–42.
- Scollon, R., 1994. As a matter of fact: The changing ideology of authorship and responsibility in discourse. *World Englishes*, 13(1), pp.33–46.
- Scollon, R., 1995. Plagiarism and ideology: Identity in intercultural discourse. *Language in Society*, 24, pp. 1–28.
- Sousa-Silva, R., Grant, T. & Maia, B., 2010. “I didn’t mean to steal someone else’s words!”: A Forensic Linguistic Approach to Detecting Intentional Plagiarism. In 4th International Plagiarism Conference: “Towards an Authentic Future” – 2010 Conference Proceedings & Abstracts. PlagiarismAdvice.org.
- Thompson, C., 2002. Discourses on plagiarism: to discipline and punish or to teach and learn? In Australian New Zealand Communication Association (ANZCA) Conference, Queensland, Australia, 10–12 July 2002. Hosted by Bond University, Queensland.
- Turell, M.T., 2008. Plagiarism. In *Dimensions of Forensic Linguistics*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp. 265–299.
- Woolls, D., 2010. Computational Forensic Linguistics: Searching for similarity in large specialised corpora. In M. Coulthard & A. Johnson, eds. *The Routledge Handbook of Forensic Linguistics*. Milton Park, Abingdon, Oxon; New York, NY: Routledge, pp. 576–590.

Acknowledgements

This work was partially supported by grant SFRH/BD/47890/2008 FCT-Portugal, co-financed by POPH/FSE.

Copyright © 2013 Rui Sousa-Silva: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

PAPERS—SECTION III
NATIONAL AND LOCAL RESPONSES TO PLAGIARISM

CULTURAL DIFFERENCES REGARDING EXPECTED UTILITIES AND COSTS OF PLAGIARISM BETWEEN HIGH TRUST- AND LOW-TRUST-SOCIETIES—PRELIMINARY RESULTS OF AN INTERNATIONAL SURVEY STUDY

Eckhard Burkatzki, Joost Platje, Wolfgang Gerstlberger

Abstract: Plagiarism is not a crime of passion but of calculation, where utility and costs are compared. The study presented asks for cultural variations regarding the perceived utilities and costs of plagiarism, focussing on the following questions:¹ (1) Do students from countries with a different level of generalized trust vary with respect to the average frequency of plagiarism? (2) Do students from countries with a level of generalized trust vary with respect to the on average perceived cost-utility structure of plagiarizing? (3) Is it possible to explain observed country-specific differences in the aggregated frequency of plagiarism by different country-specific perceptions of the cost-utility-structure of plagiarizing? Empirical research from universities in Poland, Germany and Denmark is presented. Methodically the investigation is based on bivariate statistics, Principle Component Analyses and Multiple Regression. Significant differences are found regarding both the frequency of plagiarism and the willingness to plagiarize between high-trust-, medium-trust- and low-trust-populations. In the high-trust case perceived moral costs of plagiarizing activity provides stronger disincentives for committing plagiarism than possible external sanctions. Above this, a reverse effect of administrative sanctions within high-trust- and low-trust-populations could be observed, with the remarkable recognition that in high-trust-populations the threat of administrative sanctions does not prevent but provoke deviance. Evidence has been found that differences in the aggregated frequency of plagiarism between different populations can be explained by the activity of utility-oriented actors under circumstances of a varying level of generalized trust influencing the perceived moral costs of plagiarism.

Introduction

Media reports on plagiarism in science, leading not only to the cancellation of PhD titles but also to the end of the political careers of, for example, the former German minister of defence Karl-Theodor zu Guttenberg or the former Hungarian president Pal Schmitt, show that a serious discussion on good scientific practice is necessary. When such opportunistic behaviour appears at the highest levels in society, it can be expected that this is only the tip of the iceberg. In this context, it is relevant to develop a system of self-assurance on the rules concerning scientific work and how to prevent and deal with misconduct (compare DHV/FT/AFT 2012).

In this paper, the question is addressed what are the causes of plagiarism in the scientific system, with a focus on the role of cultural factors expressed by social capital in the form of generalized trust. This trust may be an explanatory variable of differences

¹An more extensive discussion on the questionnaire results can also be found in Burkatzki et al.,(2012).

in the frequency of plagiarism in different countries, by influencing the perceived costs and benefits of plagiarism. These issues were investigated in 2010 using a questionnaire to research the views of Danish, Polish and German students.

In order to deal with these issues, first the challenge is defining plagiarism as an object of scientific research and there is discussion of the current state of research on the topic. Then, the theoretical framework for the empirical research conducted is presented. Afterwards, the methods and data used are presented, the results are discussed and analysed, and conclusions are drawn.

The research problem

The deviant character of plagiarism in science embraces the following elements (Sattler, 2007, 30ff.): the intentional use or adaption of words, ideas and/or arguments which come from an external source, are literally copied or poorly paraphrased, while sources are not mentioned and authorship is claimed by the plagiarator. In other words, plagiarism appears when protected property rights are used, and wilful ownership of the words, ideas and/or arguments is claimed (compare Loewenstein, 2003). In this connection it is implicitly assumed that the plagiarizing student or scientist is aware of this fact, and that consciousness implies full knowledge on the rules of the game of scientific practice.²

Plagiarism is a form of opportunistic behaviour, which appears due to, among other things, the existence of asymmetric information, where the supervisor or reviewer faces high transaction costs of monitoring the quality of scientific work. Three leading intentions for such malpractice among students can be reconstructed. First of all, plagiarism is employed for increasing the subjective probability of passing academic examinations. Second, the goal may be a higher grade than normally would be achieved. Third, gains of time and energy are achieved, especially when students feel not to be identified with a topic they're working on. Since the goods intended by using plagiarizing techniques are not affluently given, plagiarism can be understood as some strategic behaviour of reaching these goods by illegitimate means under conditions of scarcity. Especially this is the reason why plagiarism by some authors of white-collar crime text books is likewise subsumed as a variety of white-collar crime (Friedrichs, 2010; Payne, 2012). Taking additionally into account that students within their academic studies, implicitly or explicitly, are competing for better grades or a better starting position on the labour market, plagiarism can likewise be characterized as a crime of competition.

The main causes of plagiarism identified in research can be analysed with help of Donald Cressey's (1971) so-called fraud triangle (Fig. 1), that has been especially useful for analyzing white-collar crime. There should be an incentive or plight, which strengthens when the expected payoff increases, there should be an opportunity, this means the accessibility of illegitimate means for reaching one's goals with low risk of detection, and a neutralization factor giving justification for such behaviour.

²However, an issue for separate research is whether especially students do have the knowledge and skills to apply the rules and instruments for good scientific practice, and whether there are cultural differences in the interpretation of good scientific practice or not.

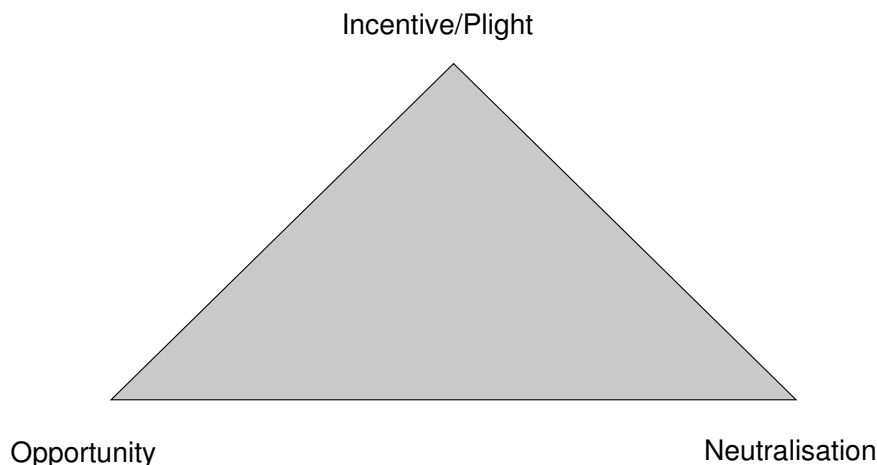


Figure 1. Fraud triangle according to Donald Cressey

Since the midst of the 1990s, the Internet has become a huge market for providing opportunities for committing plagiarism (Park, 2003; Smydra, 2004; McCabe, 2005). Above this research literature names a range of plights that are increasing the incentive to make use of plagiarism. Temporal plights are especially experienced by students spending large amounts of time for extra-curricular activities (McCabe and Trevino, 1997), practicing an undisciplined work ethics (Eisenberger and Shank, 1985), or suffering from procrastination (Patzek et al., 2012). Negative incentives for using plagiarism as a means for writing academic papers are according to the literature a high perceived severity of penalties (Bouwers, 1964; Sattler, 2007), and peer-disapproval of cheating (McCabe and Trevino, 1997). Above this a good grade and high amounts of earned credit points are likewise named a negative incentive for student cheating since the higher the capital students have accumulated during their studies the higher is the risk to be expected when loosing this capital because of detected plagiarism. Justification for plagiarism is related to the Internet as an open access for knowledge (Gajadhar, 1998; Boehm and Taggett, 2005) as well as the argument that others (peers) also plagiarize (Bouwers, 1964), which can easily lead to a degeneration of scientific culture. As a deficit of justification that might be perceived likewise as a negative incentive for plagiarism Sattler names an anticipated bad conscience that could occur when reflecting upon the option to plagiarize (Sattler, 2007).

In much research the authors seem to assume implicitly that the determinants of plagiarism identified are universally valid. However, in accordance with what is commonly assumed in institutional economic theory (see Platje, 2011), the determinants influence the willingness to plagiarize with different strengths as a consequence of cultural differences.

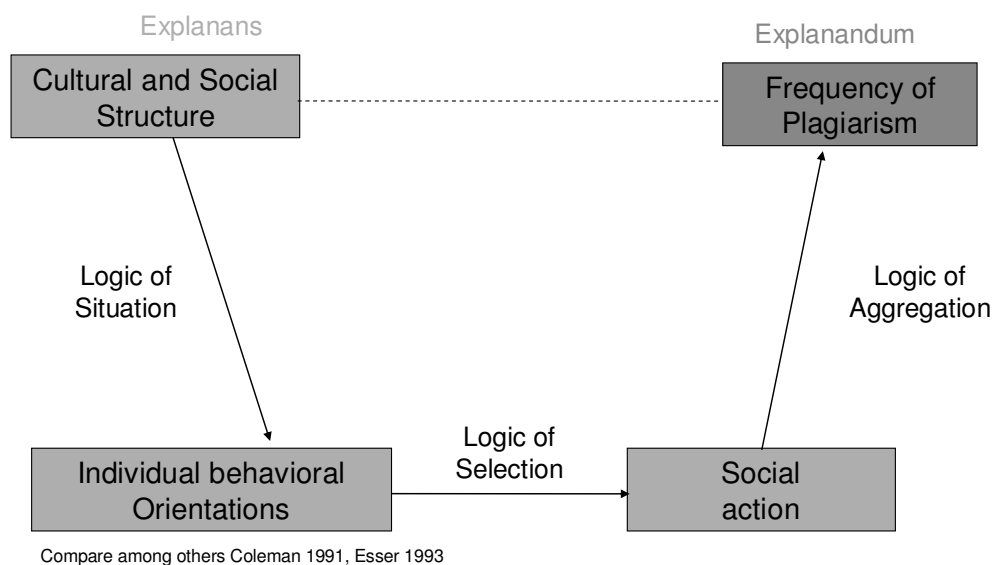


Figure 2. Model of sociological explanation

Theoretical Framework and Research Questions

Methodological individualism is the starting point of the study, assuming that phenomena arise due to utility maximizing behaviour of individuals within a social context. In addition, it is assumed that the cultural environment as well as internalized values influence the individual utility function. These assumptions are the basis of the model of sociological explanation presented in Fig. 2, making it possible to explain differences in rates of plagiarizing in different countries with different social and cultural structures.

With help of the model, micro and macro interrelations regarding social actions can be analysed. Three logical steps for the reconstruction of the emergence of social phenomena can be distinguished: the logic of situation (how do given social and cultural structures influence individual behavioural orientations as well as the individual definition of the situation), the logic of selection (how specific rules and criteria about how behavioural orientations and a given situational definition determine individual choice and actions), and the logic of aggregation (how new structural phenomena arise from the aggregation of individual actions). The first two are used for constructing the research problem.

Logic of situation

The average level of general trust in a society may be an important determinant of behavioural orientations and choices. Following Fukuyama (1995), we will distinguish between high- and low-trust societies. Fukuyama's definition of trust, the expectation of rule conforming, honest and cooperative behaviour by others whom

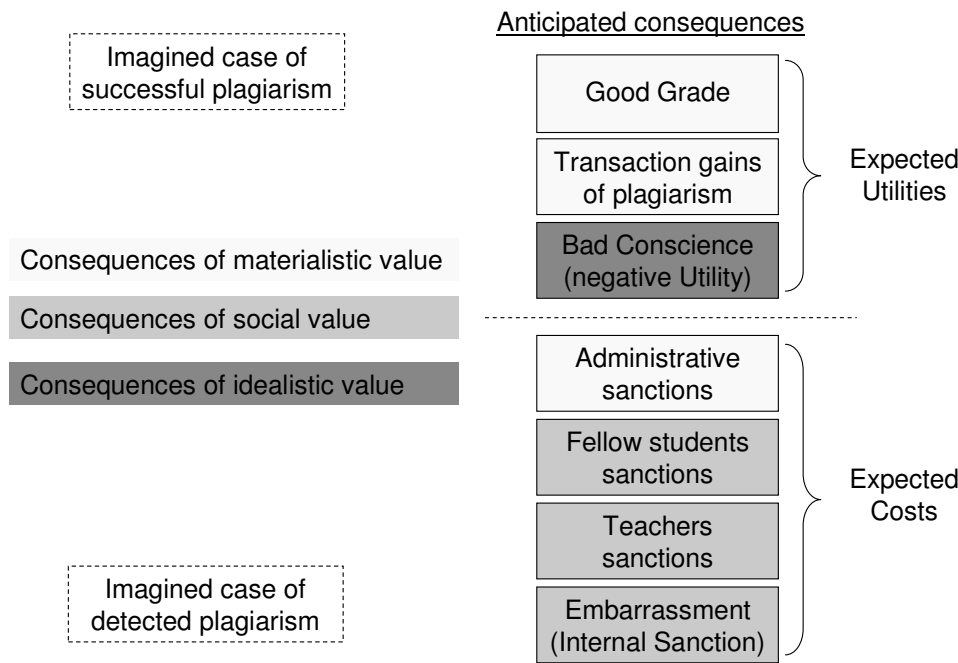


Figure 3. Model of decision making students are confronted with when thinking upon plagiarism

we not necessarily know personally, is used as the basis for the aims of the present study. It has been argued that there is a kind of self-fulfilling prophecy that trust by principals is rewarded by trustworthy behaviour by agents. Agents confronted with distrusting principals, tend to reduce their efforts to a minimum level (Falk and Kosfeld, 2006). Correspondingly, we assume that the level of generalized trust among students influences their behaviour concerning plagiarism.

Logic of selection

The fundament for the logic of selection applied in the research is rational choice theory (see Hill, 2002; Gilboa, 2010). Plagiarism is considered a risky venture, depending on the expected utility of success and expected costs of detection (see Fig. 3). Expected gains may have a materialistic nature (good grade and expected transaction gains) and an ideational nature (good or bad conscience in case of successful plagiarism). Expected costs of being detected can be of a materialistic nature (expected administrative sanctions) and social nature (expected informal sanctions by fellow students, formal and informal sanctions by academic teachers as well as embarrassment). It is assumed that the higher the positive utility and the lower the perceived costs of plagiarism, the higher the probability of individuals to engage in plagiarizing behaviour.

The theoretical frame used in this study, adapted from Fig. 2, is presented in Fig. 4. It is assumed that differences between countries regarding the frequency of plagiarism

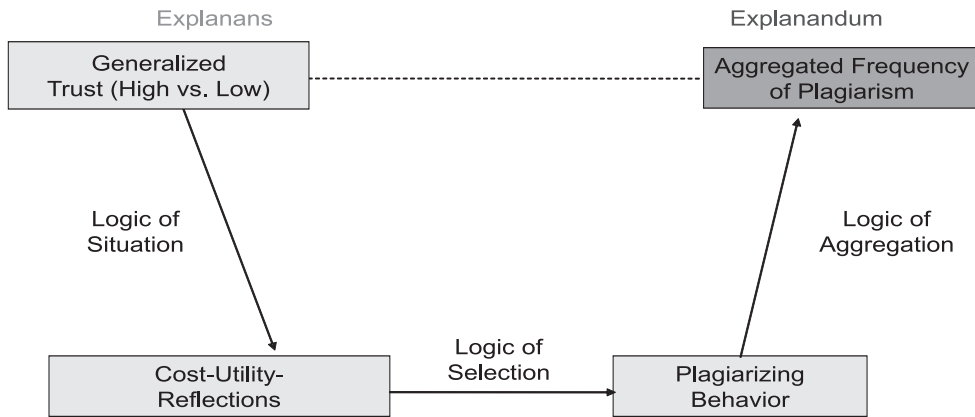


Figure 4. Theoretical frame of the study

can be explained by the level of generalized trust by way of influencing the costs and benefits of plagiarism in the form of moral costs and reduced benefits.

Research questions

Against the background of our theoretical considerations the following research questions are addressed:

- (1) Do students from countries with a different level of generalized trust vary with respect to the average frequency of plagiarism?
- (2) Do students from countries with a different level of generalized trust vary with respect to the on average perceived cost-utility structure of plagiarizing?
- (3) Is it possible to explain observed country-specific differences in the aggregated frequency of plagiarism by different country-specific perceptions of the cost-utility-structure of plagiarizing?

Method and data

Study design

In order to answer the research question, a standardized-written survey was carried out in 2010 among students of Danish, German and Polish Universities. The original version was developed in English, translated in the national language of the three countries, and afterwards translated back into English and compared with the original English version.

Method and data

The survey was carried out at the University for Applied Sciences Nordhausen in Nordhausen (Germany), the International University of Logistics and Transport, a

Table 1

Number of valid cases (n) at places of data collection

University	Number of valid cases (n)
Nordhausen (D)	152
Wroclaw (PL)	360
Opole (PL)	469
Odense (DK)	128
Total	1115

private university in Wroclaw (Poland), Opole University, a public university in the Silesian town Opole (Poland) and the Syddansk University of Southern Denmark in Odense (Denmark). The numbers of valid responses are presented in Table 1.

As only the population of particular universities was surveyed, the results are not representative for the Danish, German and Polish student population. As the socio-demographic structure of the samples differs, the data have been re-weighted according to a similar sample size and socio-demographic distribution pattern. For re-weighting as relevant variables gender (male vs. female), academic semester (3 to 6 vs. 7 and more), and mode of study (full time vs. extra occupational) of the German sample have been considered. As a consequence of the weighing procedure the student sample of each university has been of same size and according to the chosen variables of the same socio-demographic structure.

Operationalization of the research

Poland, Germany and Denmark were chosen for the research due to the differing average level of generalized trust. According to research of the European Value Study (EVS, 2008), Poland can be considered to be a low-trust society, Denmark a high-trust society while Germany being ranked in between. These results are confirmed by the results of a control question in the questionnaire. In the survey, respondents were asked to imagine to have submitted a paper they knowingly had plagiarized, which had not been detected.

Students were asked to consider how they would value the consequences of the imagined deed, as well as the estimated probability of a consequence becoming real. A second question aimed at measuring the expected costs of plagiarism by asking to imagine a plagiarized paper was submitted, which had been detected. Also here the respondents were asked to consider how they would perceive the consequences of the imagined deed, as well as the estimated probability of a consequence becoming real.

To compute the expected utilities and costs of respondents for each consequence named in the question subjective values and probabilities have been multiplied with each other. For single utilities and costs in second step have been summed up for all consequences that have been assigned to a given consequential type (compare figure 3). Last not least the summed-up values for each consequential type have been divided by the number of consequential items that beforehand were assigned to this type.

Results

Research question 1: Trust-related differences regarding plagiarism

Following the theoretical frame of the research in Fig. 4, this research question concerns the macro-correlation between generalized trust and the aggregated frequency of plagiarism. As is shown in Fig. 5, there are differences in the willingness to plagiarize between universities, with the highest willingness among the International University for Logistics and Transport in Wroclaw (Poland) and the lowest for the university in Odense (Denmark). The value of eta-square indicates that by distinguishing the places of investigation 11.3 per cent of the variance regarding the individual willingness to plagiarize within the total sample could be explained.

The results regarding the average willingness to plagiarize are confirmed by the level of self-reported plagiarism (Fig. 6), which supports the assumption that students having already plagiarized are likely to engage more in plagiarism in the future. Also in this case, the Polish universities show the highest level of plagiarism (Wroclaw—30.5%; Opole—21%). While now Nordhausen (Germany) obtains the lowest score, the high level of “no comment” may give some thoughts. These students may distrust the promise of the researchers concerning anonymity. As a consequence, the students from Odense, none of them indicating “no comment”, tend to show the lowest level of plagiarism, while the number for the Polish and German universities may be higher than reported.

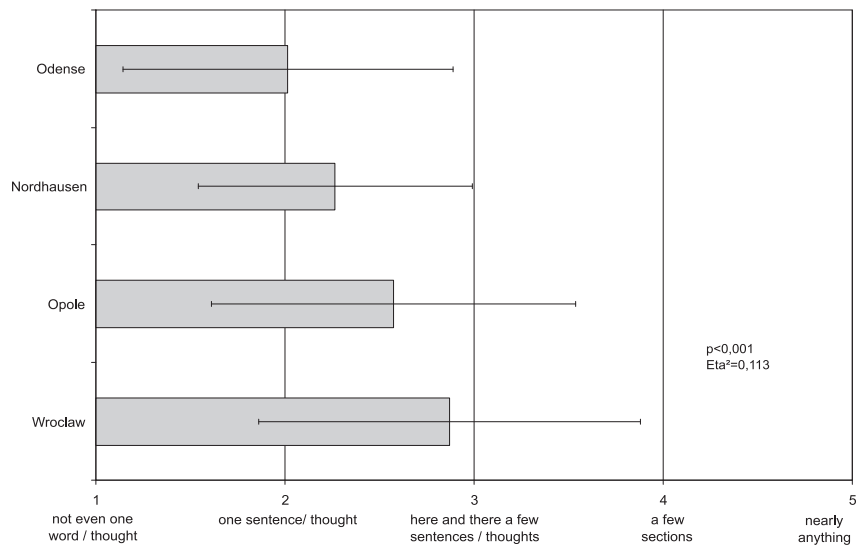
As is shown in Fig. 7, a strong correlation between generalized trust and the willingness to plagiarize can be observed. While the relationship seems to be perfectly linear, the assumption of linearity is only supported by three bivariate measures. Furthermore, on a micro-analytic level both measures are not strongly correlated, but are only associated with the inferential status of a so-called significant tendency ($p=0.09$). Nevertheless, the results are similar to the relation found between generalized trust and the perception of corruption in these countries (the higher the generalized trust, the lower the perceived corruption) (compare Fig. 8).

Research question 2: Differences of perceived utilities and costs of plagiarism in high-trust- and low-trust-samples

Regarding the research question on differences regarding perceived utilities and costs of plagiarism (micro-analytical level), focus is on the issue whether there is a correlation with the level of generalized trust (macro-level). Starting point for estimating relevant cost- and utility dimensions for committing plagiarism is the regression model depicted in Fig. 9.

To reduce complexity with respect to relevant dimensions of the perceived cost-utility-structure of plagiarism only those factors are taken into account that have a statistical significant effect on the individual willingness to plagiarize. Symbolically distinguished within the figures are the following levels of statistical significance: *** $p \leq 0,001$, ** $p \leq 0,01$, * $p \leq 0,05$ and † $p \leq 0,10$ (compare Bortz, 1999). Since the sample data analysed within this study do not in a strong sense have a representative status, documented information regarding the statistical significance of correlations

How much thoughts and citations of external sources would you be maximally willing to adopt in a seminar paper without indicating the origins?



Data: Studying and Ethics, 2010
Source: Own computations

Note: The horizontal lines connected to the bars indicate the simple standard deviation of values around the sample-specific mean value.

Figure 5. Average willingness to plagiarize within student samples

cannot be interpreted in an inferential manner and do only have a descriptive character. The results of regression analyses for the student samples first of the Polish Opole, second of the Danish Odense and third of the German Nordhausen are depicted in figures 10.1 to 10.3. The regression results for the Wroclawian student sample have not been documented since strong problems of co-linearity (VIF-value > 4) appeared within this model.

From the results it becomes clear that cost-utility reflections are quite important for the individual willingness to plagiarize. However, regression results vary depending on the average level of generalized trust. First, it seems that moral self-commitment (bad conscience) is more relevant in a high-trust setting (Odense) than a medium-trust setting (Nordhausen) and in particular a low-trust setting (Opole). Thus, the expectation of honest behaviour by others seems to positively influence the controlling and steering effect of morality on choices.

Second, administrative sanctions seem to have a stronger preventive effect in the low-trust case (Opole), while in the high-trust case such sanctions seem to cause an increasing willingness to plagiarize (Odense). This may be explained by the theory of violated psychological contracts (Litzky et al., 2006). The idea is that students, being strongly committed to norms of academic integrity, are suffering conscience-based pains when thinking upon the behavioural option to plagiarize. When faced

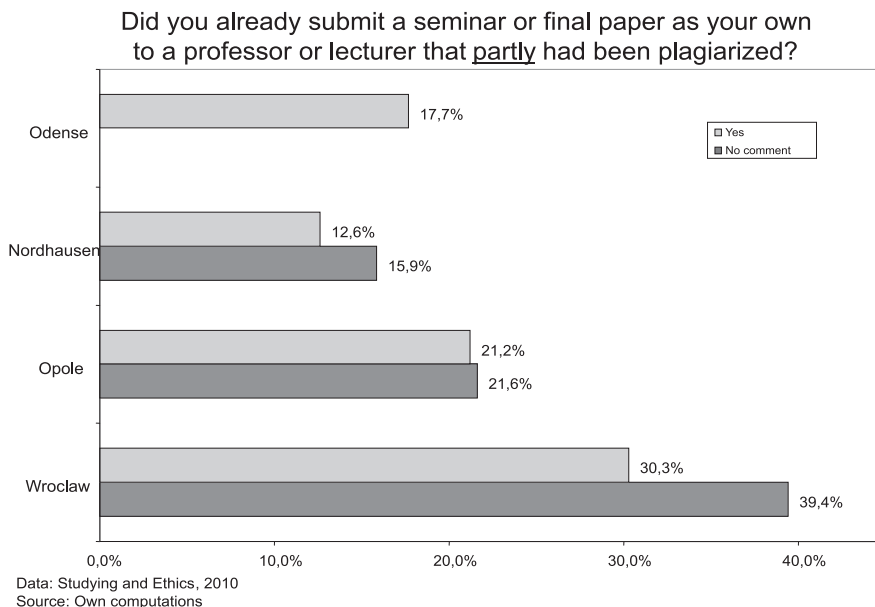


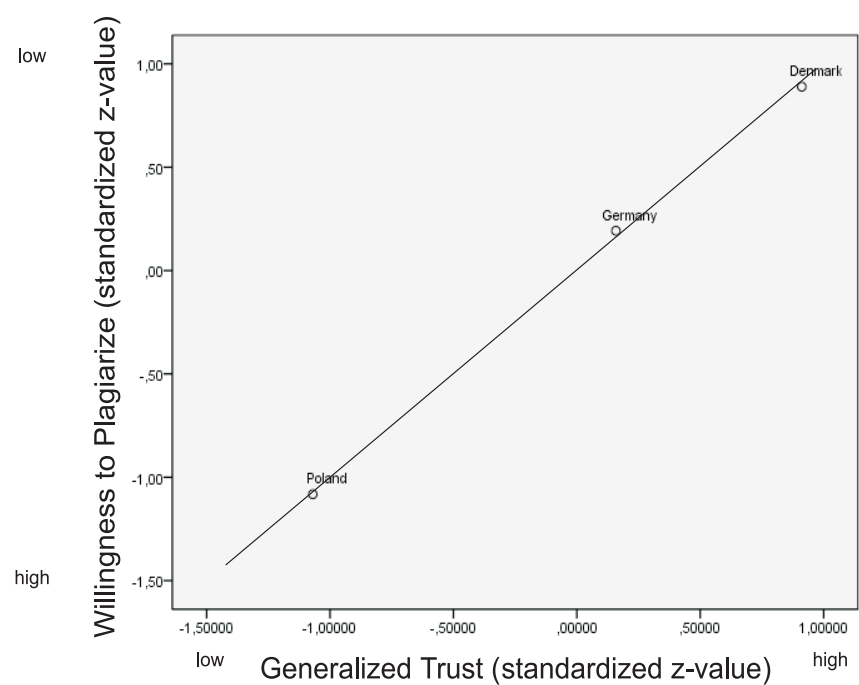
Figure 6. Share of students reporting plagiarizing activities within student samples

by the threat of administrative sanctions they may feel that an implicit contract of reciprocal psychological appreciation is violated. Those negative emotions are promoting a behaviour (plagiarism) challenging especially those sanction systems they feel threatened by.

Thirdly, in the Opole low-trust sample, the mentioned administrative sanctions, but also informal sanctions from fellow students inhibit plagiarism. Moral discomfort (bad conscience) and feelings of embarrassment when detected seem not to be a discouraging factor. However, the expected costs of teachers' sanctions are likewise enhancing students' individual willingness to plagiarize. An explanation may be that in case of lack of authority of a teacher threatening with hard sanctions, students may develop an attitude of reactancy (see Brehm and Behm, 1981). Another explanation may be a "crime as play" (Richards et al., 1979) scenario, where students knowing the probability of being caught is low, play a kind of game with teacher showing the threats do not hurt them. This culturally embedded dissidence against an established, but questioned, order or authority may have its roots in Polish long experience of foreign domination (compare Morawska, 1984).

Research question 3: Explaining sample-related variances of the average willingness to plagiarize with respect to trust-related differences regarding expected utilities and costs of plagiarism

The third research question focuses on the perceived costs and utilities of plagiarism as an explanatory variable of the willingness to plagiarize. Furthermore, it is researched



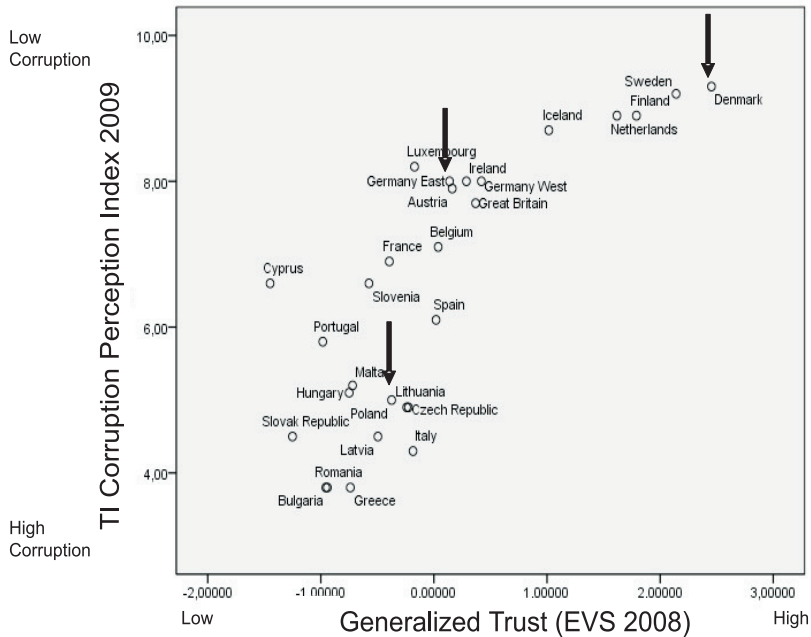
Data: Studying and Ethics, 2010
Source: Own computations

Figure 7. Willingness to plagiarize and generalized trust within the Polish, German, and Danish student sample

whether the costs and utilities themselves are influenced by the level of generalized trust. These issues are analysed with use of multiple linear regression analysis in combination with the construction of interaction effects..

While three samples are included in the analysis, only two causal paths are considered for the reason that the three sample-affiliations together constitute somehow a decomposed ordinal scale of generalized trust, wherein the lowest manifestation, in our case affiliation to the Polish student samples, has to serve as a reference category. We first tried to calculate the independent effect of the level of generalized trust on the individual willingness to plagiarize. Then it was investigated whether the influence of the sample-affiliation as an aggregated variable on the willingness to plagiarize within the analytical frame of multivariate regression analysis loses significance when we additionally include individual cost-utility-reflections into the statistical model. If this were the case, this would support the assumption of methodological individualism that the origin of macro-phenomena can be explained by the utility-oriented actions of individual actors.

Results of the regression analyses are documented in figures 11.1 and 11.2. The standardized regression coefficients shown in Fig. 11.1 confirm the earlier presented findings



Data: European Value Study 2008 (z-standardized data; country specific mean values), CPI 2009
Source: Own computations

Figure 8. Willingness to plagiarize and generalized trust within the Polish, German, and Danish student sample

that the level of generalized trust influences the average willingness to plagiarize. The higher the level of such trust, the lower is the willingness to plagiarize. This effect is stronger for the Danish sample, featured by a higher level of generalized trust. A share of 9.9 per cent of the observed variance regarding the willingness to plagiarize can be explained by the aggregated level of trust within the students' environment.

Analysing expected utilities and costs of plagiarism into the regression model (Fig. 11.2), only bad conscience and administrative sanctions seem to influence the willingness to plagiarize. Under control of behavioural orientations (anticipated utilities and costs) the influence of generalized trust on willingness to plagiarize within the given multivariate model decreases.

Generalized trust and behavioural expectations within this multivariate regression model explain 22.4 per cent of the variance of the dependent variable. However, although cost and benefit considerations decrease the strength of influence of generalized trust, which is visible in the German medium-trust case, in the Danish high-trust sample, within the multivariate main effect model, it remains an independent predictor of plagiarism. This indicates, corresponding to basic assumptions of methodological individualism, that macro-influences of aggregated generalized trust within students' environment are partly mediated by the behavioural expectations.

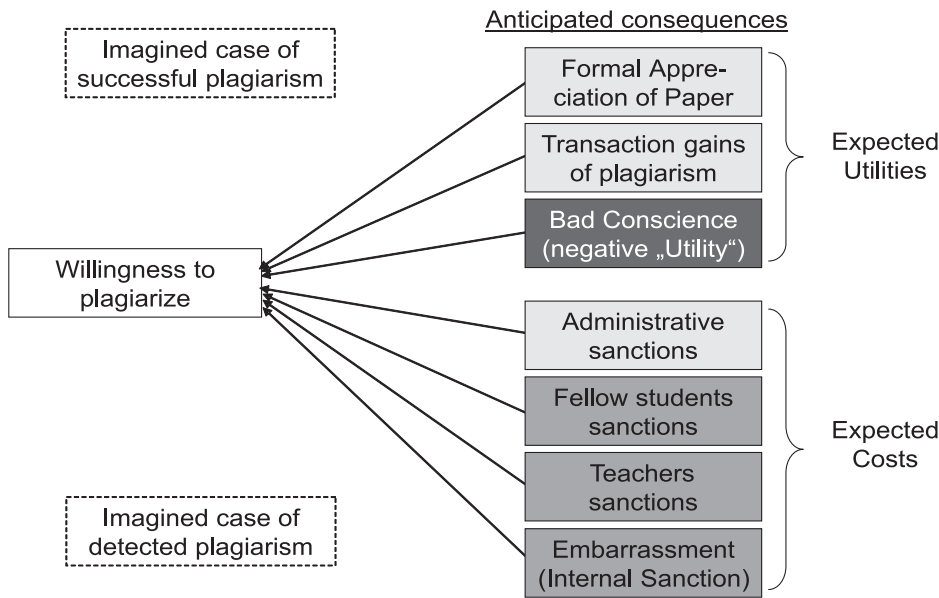


Figure 9. Regression model for comparing relevant dimensions of the perceived cost-utility-structure of plagiarism between high-trust- and low-trust-student-samples

When trying to explain why the macro influence of generalized trust on the willingness to plagiarize cannot be explained by the behavioural expectations of actors, it should be kept in mind that one variable, the expectation of administrative sanctions, reversely influences plagiarism within high- and low-trusts settings. Such a reverse and context-dependent effect was not included in the main model of regression used in earlier analysis. In order to do this, the effect has to be included into the linear additive regression model multiplicative terms, allowing for the reconstruction of interaction effects between independent variables (compare Jaccard and Turrisi, 2003). This allows for statistical analysis whether behavioural expectations partial out the independent macro effect of generalized trust on the willingness to plagiarize. If this would be the case, this would support the considerations discussed earlier that not only, like methodological individualism postulates, phenomena such as plagiarism, crimes, fashions and economic crises can be explained by utility maximizing behaviour of individual actors within a social context, but that also cultural factors and internalized values provide incentives for individual action.

The estimation results of the linear regression are presented in Table 2,³ showing the standardized regression coefficients of all regression models that have been estimated with data of the total sample. These data allow for analysing the question whether

³As suggested by Allison (1998, p. 168), while including interaction effects within a given regression model, the main effects of all relevant variables are kept within the model even when they are producing problems of multicollinearity (indicated e.g. by variance inflation factors > 4). Otherwise regression estimations for the interaction terms could be strongly biased. However Allison likewise stresses that coefficients for main effects within regression models that likewise contain interaction effects should not be interpreted.

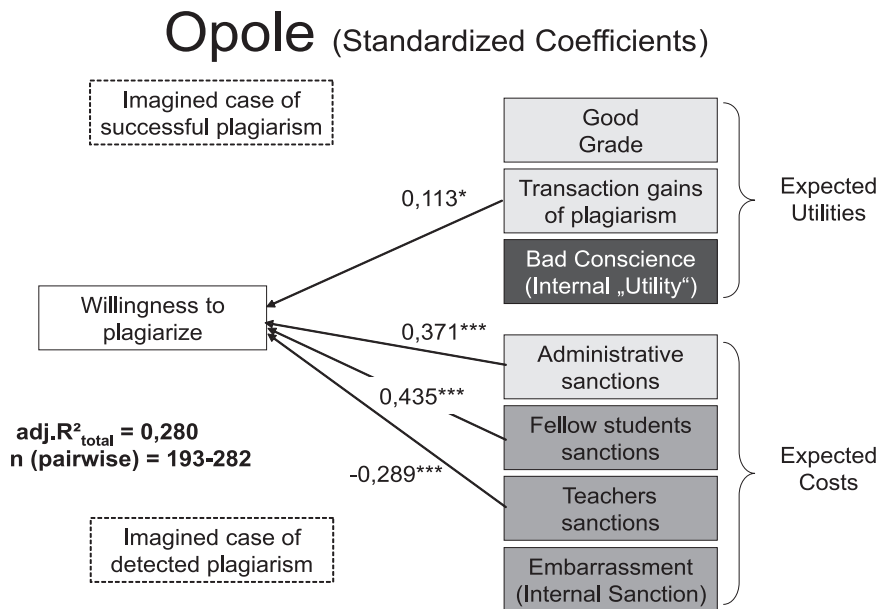


Figure 10.1. Regression of willingness to plagiarize on perceived utilities and costs of plagiarism among Opolian students (low-trust-sample)

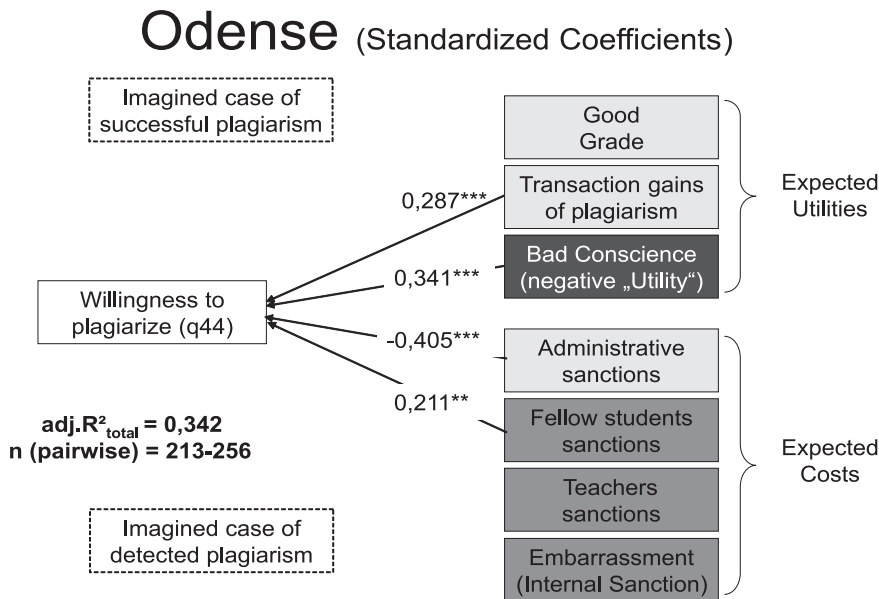


Figure 10.2. Regression of willingness to plagiarize on perceived utilities and costs of plagiarism within the Odensian student sample (medium-trust-sample)

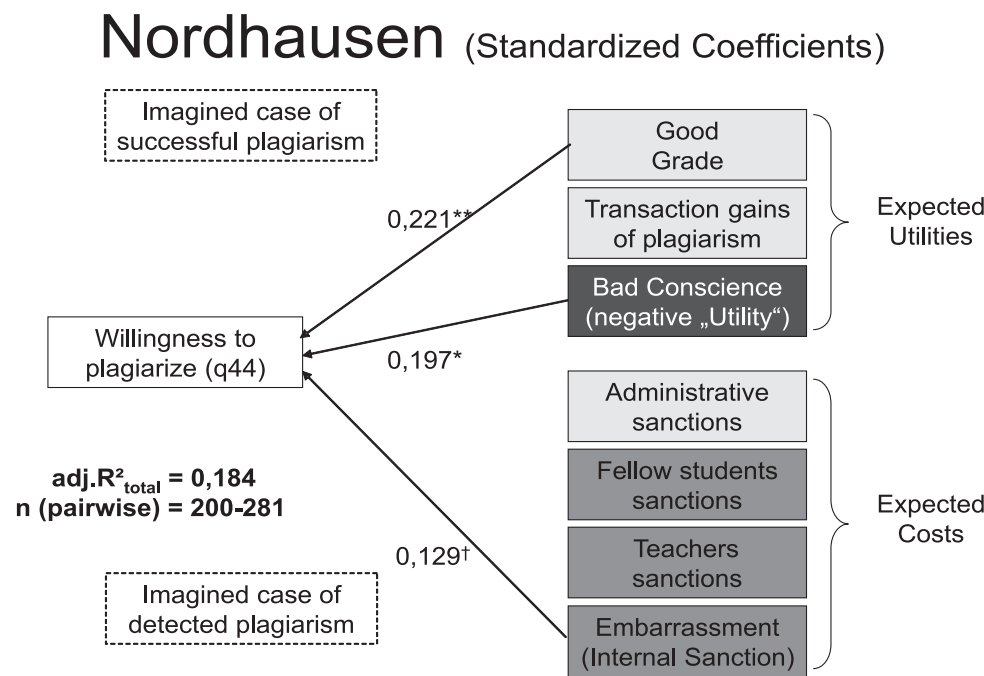


Figure 10.3. Regression of willingness to plagiarize on perceived utilities and costs of plagiarism within the Nordhausen student sample (high-trust-sample)

the macro level effects of generalized trust lose statistical significance or not when controlling for both: the main effects of behavioural expectations on the micro-level of investigation and the macro-micro-interactions described above. It turns out that the macro-level effects lose significance when interaction effects both for utility- and cost-related expectations regarding the consequences of plagiarism are included. This confirms our assumption that the generalized level of trust within a population, besides situational incentives, is an important contextual factor for behavioural orientations of actors and for individual action.

However, the regression results confirming this assumption have been estimated with the so-called listwise-mode of analysis, denoting a mode of multivariate analysis excluding all cases from statistical estimation where just on one of the variables included in the model has a missing value. Although in literature this method is strongly recommended, a problem is that its application sometimes involves a great loss of cases to be analysed, in particular when variables are included into the multivariate model which themselves exhibit a great number of missing values (see Reinecke, 2005). Thus, the sample size used for estimating the model coefficients is the more reduced the more variables have been included in the model. As a consequence, the sample size that could be used to estimate the coefficients is not even half as large as the initial sample size. Since the significance of coefficients strongly depends on sample size and since of course the size of coefficients strongly depends on variation of data

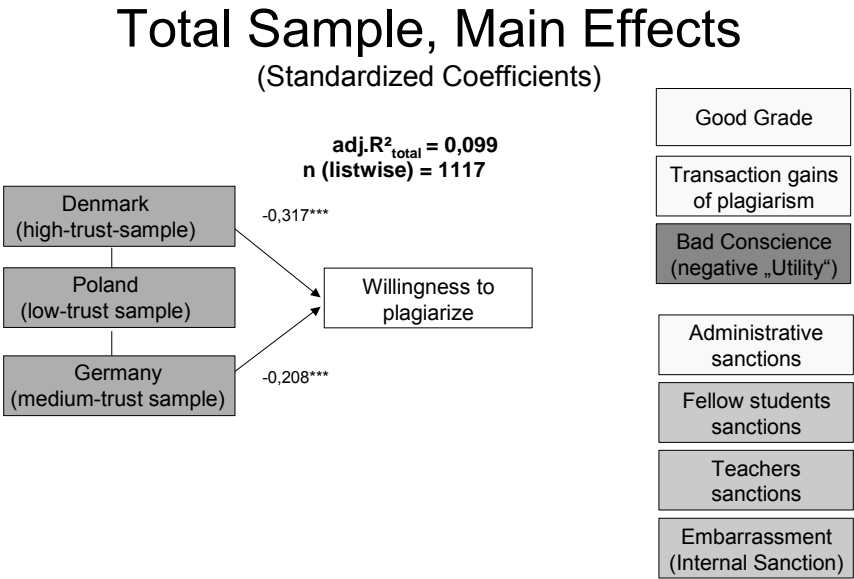


Figure 11.1. Regression of willingness to plagiarize on generalized trust in respondents environment, measured on an aggregate level by respondents affiliation to a country-specific student sample

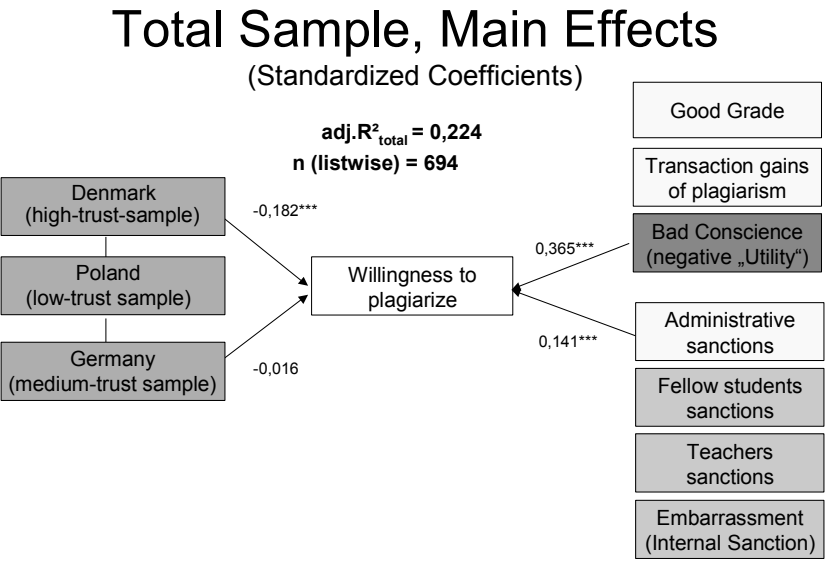


Figure 11.2. Regression of willingness to plagiarize on generalized trust in respondents environment, measured on an aggregate level by respondents affiliation to a country-specific student sample, and on individual utility-costs-reflections upon possible consequences of plagiarism

and correspondingly on the composition of a given sample, there exists the risk that the findings documented above may be a statistical artefact.

Conclusion

In this paper, cultural variations regarding the perceived utilities and costs of plagiarism have been the topic of investigation. Survey results were presented in order to deal with the three research questions focused on the influence of the level of generalized trust in the average willingness to plagiarize, the influence of generalized trust on the average perceived costs and utilities of plagiarism and the influence of country-specific perceptions of the costs and utilities on the aggregated frequency of plagiarism.

The results of the survey indicate that there exist significant differences regarding both the frequency of plagiarism and the willingness to plagiarize between high-trust-, medium-trust- and low-trust-populations. Furthermore, the higher the level of generalized trust, the lower is the willingness to plagiarize. In particular in the high-trust case, moral costs of plagiarism seem to increase, reducing the willingness to plagiarize. There also seems to be a reverse effect of administrative sanctions within high-trust- and low-trust-populations. It is remarkable that in the high-trust case the threat of administrative sanctions does not prevent but provoke deviance. The opposite is true for low-trust-populations, where administrative sanctions have a preventing effect on plagiarism. A third result from regression analysis is that differences in the willingness to plagiarize can be explained by the activity of utility-oriented actors under circumstances of varying social and cultural structure. The level of generalized trust seems to have a strong influence on the perceived moral costs of plagiarism, preventing the appearance of such opportunistic behaviour.

However, the results of the research have to be interpreted with care. First of all, the comparisons made are based on case studies in Poland, Germany and Denmark. Thus, the outcomes cannot be extrapolated for the whole countries' student population, and comparison with data from other or new surveys is necessary. Second, regarding selected survey questions, non-response tended to be high, while missing data seeming not to be distributed randomly within the student samples. The listwise mode used to deal with this problem may lead to the risk of finding a statistical artefact. More research is required to confirm or reject the findings. Third, in particular German and Polish students did not answer questions regarding the observed frequency and the administrative treatment of plagiarism. Thus, differences found regarding generalized trust may not be the result of country-specific cultural structures, but by the organizational environment. To deal with this issue, research within randomly selected universities in the countries in question is required. Fourth, due to the limited number of universities researched as a consequence of economic limitations, multilevel analysis in a strong sense could not be conducted, which could lead to a bias on the coefficient estimates.

Although there are limitations to the study, some conclusions can be drawn which may be useful for the creation of policy for preventing plagiarism. First, administrative sanctions may not be such an effective mean in fighting plagiarism. Especially in a high-trust setting, they even may lead to counterproductive behaviour. Second,

Table 2

Linear regression (listwise) of willingness to plagiarize on main effects of generalized trust and behavioural expectations and interaction effects of behavioural expectations with generalized trust

Listwise	Models (standardized coefficients)								
	1	2	3	4	5	6	7	8	9
Generalized trust									
High-trust-sample (Dnmk)	-0,317***	-0,161***	-0,146***	-0,209***	-0,146***	-0,149**	-0,182***	-0,052	-0,075
Medium-trust-sample (Ger)	-0,208***	-0,038	-0,085*	-0,081*	-0,07†	0,003	-0,016	-0,038	-0,037
Low-trust-sample (PL) (Reference)	–	–	–	–	–	–	–	–	–
Exp. utilities, Main Effects									
Good Grade		-0,014	-0,055			0,036		0,143*	0,154***
Transaction Gains		0,014	-0,219***V			-0,003		-0,246***V	-0,245*** V
Good/Bad Conscience		0,387***	0,570***			0,307***	0,355***	0,447***V	0,444*** V
Exp. costs, Main Effects									
Administrative Sanction				0,090*	0,354***	0,125**	0,141***	0,520***	0,533***
Fellow Students Sanction				0,165***	0,304***V	0,096†		0,133 V	0,141**
Teachers Sanctions				-0,033	-0,085 V	-0,059		-0,051 V	-0,092
Embarrassment				0,215***	0,014 V	0,018		-0,356***V	-0,342***V
IAKT: Utilities*Country									
Good Grade, Dnmk			0,063					0,004	–
Good Grade, Germany			0,110*					0,027	–
Transaction, Dnmk			0,237***					0,295***	0,287***
Transaction, Germany			0,178**					0,113*	0,126*
Conscience, Dnmk			-0,189***					-0,186* V	-0,190**V
Conscience, Germany			-0,182***					-0,125*	-0,139*
IAKT: Costs*Country									
Admin. Sanction, Dnmk					-0,354***			-0,450***	-0,449***
Admin. Sanct., Germany					-0,189***			-0,278***	-0,305***
Fellows Sanction, Dnmk					-0,041			0,089 V	–
Fellows Sanct., Germany					-0,143**			-0,070	–
Teachers Sanction, Dnmk					0,056			-0,082	–
Teachers Sanct., Germany					0,031			-0,007	–
Embarrassment, Dnmk					0,178**			0,302*** V	0,315***
Embarrassment, Germany					0,136*			0,226***	0,207***
ANOVA									
R ²	0,099	0,186	0,222	0,189	0,262	0,211	0,224	0,359	0,358
Significance	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001
N	1117	718	718	656	656	532	694	532	532

*** p ≤ 0,001, ** p ≤ 0,01, * p ≤ 0,05 and † p ≤ 0,10

V: Variance Inflation Factor > 4 (problem of multicollinearity)

Source: Own Computations

self-conscience seems to have a larger impact within a high-trust setting. Student self-governing like the introduction of a moral code and student honour boards may improve self-conscience (McCabe and Trevino, 1993; McCabe et al., 1999; McCabe et al., 2001), but without the existence of generalized trust, the effectiveness may be limited.

References

- Ackermann, K. (1992) Fälschung und Plagiat als Motiv in der zeitgenössischen Literatur. Heidelberg: Carl Winter Universitätsverlag.
- Allison, PD. (1998) Multiple regression: a primer. Thousand Oaks, CA: Pine Forge Press.
- Baird, JS. (1980) Current trends in college cheating. In: *Psychology in the Schools*, 17, 515–522.
- Boehm, DC. & Taggett, L. (1998) About Plagiarism, Pixels and Platitudes. Available from: <http://www.svsu.edu/dboehm/Assignments.htm> (accessed 16. Dec. 2012)
- Bortz, J. (1999) Statistik für Sozialwissenschaftler. Berlin; Heidelberg u.a.: Springer
- Bouwens, WJ. (1964) Student dishonesty and its control in college. New York: Bureau of Applied Social Research, Columbia University
- Brehm, SS. & Behm, JW. (1981) Psychological reactance. New York: Academic Press
- Burkatzki, E., Gerstlberger, W. & Platje, J. (2012) Cultural differences regarding expected utilities and costs of plagiarism—preliminary results of an international survey study. *Economic and Environmental Studies*, 12 (3), 235–279
- Coleman, JS. (1991, orig. 1990) Grundlagen der Sozialtheorie, Bd. 1: Handlungen und Handlungssysteme. München: Oldenbourg
- Cressey, Donald R. (1971, orig. 1953) Other people's money: a study of the social psychology of embezzlement. Belmont
- DHV; FT; AFT (2012) Gute wissenschaftliche Praxis für das Verfassen wissenschaftlicher Qualifikationsarbeiten. Available from: http://www.hochschulverband.de/cms1/fileadmin/redaktion/download/pdf/resolutionen/Gute_wiss._Praxis_Fakultaetentage.pdf (accessed 16. Dec. 2012)
- Dießel, M. (2012): Die Wirkung von Vertrauen und Misstrauen auf Entscheidungen in sozialen Interaktionen. Bonn: Inaugural-Dissertation (Available from: <http://hss.ulb.uni-bonn.de/2012/2771/2771.pdf>)
- Eisenberger, R. & Shank, DM. (1985) Personal work ethic and effort training affect cheating. *Journal of Personality and Social Psychology*, 49, 520–528
- Esser, H. (1993) Soziologie. Allgemeine Grundlagen. Frankfurt a. M.: Campus
- Esser, H. (1999) Soziologie: spezielle Grundlagen, Bd.1: Situationslogik und Handeln Frankfurt a. M.: Campus
- Falk, A. & Kosfeld, M. (2006) The hidden costs of control. *American Economic Review*, 96, 1011–1030
- Friedrichs, DO. (2010) Trusted Criminals. Belmont, CA: Wadsworth
- Fukuyama, F. (1995) Konfuzius und die Marktwirtschaft – Der Konflikt der Kulturen (orig.: Trust). München: Kindler
- Gajadhar, J. (1998) Issues in plagiarism for the new millenium: An Assessment Odyssey. Available from (Archival request): <http://ultibase.rmit.edu.au> [accessed 13 Dec. 2012]
- Gilboa, I. (2010) Rational Choice. Cambridge, Mass: MIT Press

- Hill, P. (2002) *Rational-Choice-Theorie*. Bielefeld: Transcript
- Jaccard, J. & Turrisi, R. (2003) *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage
- Litzky, B., Eddleston, KA. & Kidder, DL. (2006) The Good, the Bad, and the Misguided: How managers inadvertently encourage deviant behaviors. *The Academy of Management Perspectives*, 20, 1–103
- Loewenstein, U. (2003) II. Plagiat – IV. Doppelschöpfung. In: idem (Eds.): *Handbuch des Urheberrechts*. München: Beck
- Lüdemann, Ch. & Ohlemacher, Th. (2002) *Soziologie der Kriminalität. Theoretische und empirische Perspektiven*. Weinheim, München: Juventa
- McCabe, D. (2005) In new CAI Research conducted. Available from: http://www.academicintegrity.org/cai_research.asp [accessed 13 Dec. 2012]
- McCabe, DL. & Trevino, L. (1993) Academic dishonesty: Honor codes and other contextual influences. *Journal of Higher Education*, 64, 522–538
- McCabe, DL. & Treviño, L. (1997) Individual and contextual influences on academic dishonesty: A multicampus investigation. *Research in Higher Education*, 38, 379–396.
- McCabe, DL., Trevino, LK. & Butterfield, KD. (1999) Academic integrity in honor code and non-honor code environments: A qualitative investigation. *The Journal of Higher Education*, 70 (2), 211–234.
- McCabe, D. L.; Trevino, L.; Butterfield, K. D. (2001) Cheating in academic institutions: A decade of research. *Ethics and Behavior*, 11, 219–232.
- Morawska, E. (1984) Civil religion vs. State power in Poland. *Society*, 21, 29–34.
- Park, C. (2003) In other (people's) words: plagiarism by university students—literature and lessons. *Assessment & Evaluation in Higher Education*, 28, 471–488.
- Patrzek, J., Sattler, S., van Veen, F., Grunschel, C. & Fries, St. (2012) Einfluss von akademischer Prokrastination auf studentisches Betrugsverhalten: eine längsschnittliche Analyse. Presentation at the 48th Kongress der Deutschen Gesellschaft für Psychologie, Universität Bielefeld.
- Payne, BK. (2012) *White-collar crime. A text-reader*. Thousand Oaks, CA: Sage.
- Perry, AR., Kane, KM., Bernesser, KJ. & Spicker, PT. (1990) Type A behavior, competitive achievement-striving, and cheating among college students. *Psychological Reports*, 66, 459–465.
- Platje, J. (2011) *Institutional Capital: creating capacity and capabilities for sustainable development*. Opole: Opole University Press.
- Popper, K. (1977) *Die offene Gesellschaft und ihre Feinde, Bd.2: Falsche Propheten. Hegel, Marx und die Folgen*. München: Francke Verlag.
- Reinecke, J. (2005) *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg.
- Richards, P. & Berk, Richard A.; Forster, Brenda (1979) *Crime as play: delinquency in a middle class suburb*. Cambridge: Ballinger.
- Sahlins, M. (1972) On the Sociology of primitive exchange. In: Sahlins, M. (Eds.) *Stone Age Economics*. New York: De Gruyter, 185–275.
- Sattler, S. (2007) *Plagiat in Hausarbeiten*. Hamburg: Kovac.
- Sims, RL. (1993) The relationship between academic dishonesty and unethical business practices. *Journal of Education for Business*, 68, 207–211.
- Smydra, RV. (2004) The challenge of plagiarism control in universities and colleges. In: *Proceedings of the 2004 Association for Business Communication Annual Convention*.

Weber-Wulff, D. (2012) Plagiarism detection software—Silver Bullet or Waste of Time
Workshop-Presentation at the conference “Mastering the challenges in higher education: considering the way students learn, cheat and enhance performance”, 07–09 February 2012, Bielefeld, Germany.

Authors

Eckhard Burkatzki, burkatzki@ihi-zittau.de, TU Dresden—International Institute Zittau, Germany;

Joost Platje, jplatje@uni.opole.pl, Opole University, Poland;

Wolfgang Gerstlberger, University of Southern Denmark, Denmark

Copyright © 2013 Eckhard Burkatzki, Joost Platje, Wolfgang Gerstlberger: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

STUDENTS' PERCEPTIONS ON PLAGIARISM AND RELEVANT POLICIES IN CYPRUS

Angelika Kokkinaki, Melpo Iacovidou, Catherine Demoliou

Abstract: This paper discusses the findings of a research conducted among university students in Cyprus with the purpose of exploring students' awareness and perceptions relating to academic plagiarism. It is clear from the findings of this study that there is a need for a clear and uniform definition and understating of the terms "plagiarism" and "academic dishonesty". There is also a need to develop mechanisms for communicating these definitions both to students and to faculty in addition to ensuring that these terms have been understood. In the setting up of policies, procedures and penalties for plagiarism and academic dishonesty, it is necessary to have uniformity and consistency in their implementation and especially in the implementation of penalties. The results also indicate that the use of tools, such as software for prevention and/or detection of plagiarism, would definitely contribute positively in reducing plagiarism.

Introduction

Easy accessibility to online educational resources and to scientific publications has facilitated academic plagiarism. The increased number of identified cases at Higher Educational Institutions (HEIs) raises concerns among academic faculty and higher administration as well as among policy makers at quality assurance institutions. The rising concerns on the extent of plagiarism practiced by students worldwide, has promoted the research and publications on prevention, detection and deterrence of plagiarism in order to address various interrelated aspects.

Research on the contributing factors for plagiarism (Chanock 2008) and the deterrence mechanisms (Carroll 2007) have associated good practices on academic writing with anti-plagiarism approaches. A number of educational resources are available online or published which offer advice primarily to students and also to academics willing to assist their students. They outline best practices that contribute towards prevention of plagiarism. Published educational resources include principles and examples on paraphrasing, referencing and properly addressing sources. Awareness of best practices, per se, is only a part of any policy aiming to prevent plagiarism. Repetitive practice on academic writing (Emerson, Rees & MacKay, 2005) and proper examples have also been proposed in the framework of anti-plagiarism approaches.

Another relevant approach is the provision of detailed feedback to students' work (Barrett & Malcolm, 2006). This can be accomplished as part of a formative assessment of students' work, which includes software packages that support identification of text similarities (Davis 2009). Recently, Ireland and English (2011) have proposed the concept of a "safe environment" where students are allowed to "plagiarize" as a means for gradually developing proper academic writing skills. This approach was made possible as a result of the development of software tools, such as Turnitin®, which rely on detection and prevention of plagiarism and allow formative assessment (Bennett, 2005; Davis, 2007; Murray, 2006).

Combating plagiarism effectively while maintaining consistency in the standards and quality of higher education across the European Union (EU,) requires the identification of any gaps, best practices and case studies on plagiarism across the member states. Research on practices and policies on anti-plagiarism at universities has been conducted primarily at English-speaking countries (MacCabe, 2005; Hayes and Introna, 2005) and less so at non-English speaking ones (Carroll and Zetterling 2009). This paper reports on the findings of research conducted among university students in Cyprus HEIs as part of the “IPPHEAE: Impact of Policies for Plagiarism in Higher Education across Europe ERASMUS-Lifelong Learning project” (2010–2013). The purpose of this research was to explore students' awareness and perceptions relating to academic plagiarism. More specifically, the survey attempted to identify students' interpretation of the term “plagiarism”, their awareness of plagiarism, the extent of plagiarizing and the reasons for this, and the existence and implementation of policies, procedures and penalties for dealing with plagiarism at the institution of their study.

Methodology

The research was undertaken in two stages. Initially, the questionnaire was designed based on the feedback received from the five partner institutions participating in the IPPHEAE project. Prior to administration, the questionnaires were pilot-tested for reliability with the utilisation of the test re-test method and the use of experts.

The second stage was the survey of the targeted group. The target population of the study included all students, undergraduate and graduate, pursuing an academic degree, in public and private universities in Cyprus. The survey questionnaire included 33 questions, some open-ended and some with multiple answers, which were made available either on-line or as hard copies. A total of 318 correctly completed questionnaires were collected. Prizes, based on a lottery draw, were offered to students to encourage them to complete the questionnaire.

Results and Discussion

As shown in Table 1, the majority of the responders (74.7%) were undergraduate students, studying on a full-time basis (81.5%), in their last year of study (34.8%) and between 21–25 years old (55.4%).

With regard to the definition of plagiarism, the most popular answers (in terms of descending popularity) given by students were: “Copy and paste, without acknowledgement of the source you took your material from”; “Plagiarism is a close imitation or form of cheating” and “The uses of work/research of others without citation, thus making it appear as your own work”. These results indicate a lack of a clear understanding of “plagiarism” by students and are in agreement with Carroll (2007) conclusions that students' perception of the term is not shared by universities or academicians.

As shown in Fig. 1 there are no differences in the number of students who become aware of plagiarism either before or during their undergraduate studies. However, the majority of students (70.8%) appear to learn how to cite and reference only during their undergraduate studies (Fig. 2) from sources like the web, course booklets and class notes (Fig. 3). It is interesting to note that these are also the sources that students report

Table 1

Demographics: Cyprus Data

Percentage of total students responded (<i>n</i> = 318)							
Degree	%	Mode of study	%	Year of study	%	Age (years)	%
Bachelor's	74.7	Full time	81.5	First	27.2	<20	13.1
Master's	25.3	Part time	12.0	Second	17.0	21–25	55.4
PhD	0.0	Other	6.5	Third	17.4	26–30	9.6
				Fourth	34.8	31–40	16.9
				Fifth	3.3	41–50	4.8
				Sixth or higher	0.3	50+	0.3

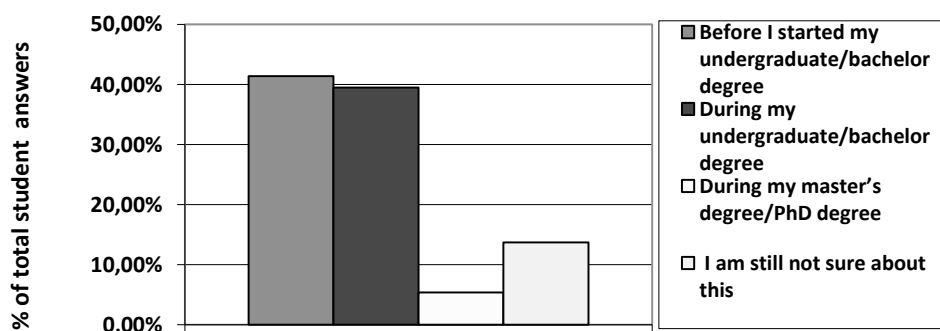


Figure 11.2. Timing of plagiarism awareness

to have contributed to their awareness of plagiarism (Fig. 4). These results indicate that the period of university undergraduate training is an important period for sensitizing students to plagiarism.

Awareness of plagiarism may be part of the students receiving adequate training on academic writing and anti-plagiarism issues (70.8% agree) as indicated from the results to the relevant question in Table 2. Although, on the basis of the questions used it was not evident how early the students receive such training, it was of interest to see that the majority of them (70%) received it from “tutors and lecturers and/or during a course/module” (60%) as shown in Table 3.

Table 2

Percentage of students' responses on received training on scholarly academic writing and anti-plagiarism issues

Strongly Disagree	Disagree	Not sure	Agree	Strongly Agree	Not applicable
2.2%	11.1%	10.5%	47%	23.8%	5.4%

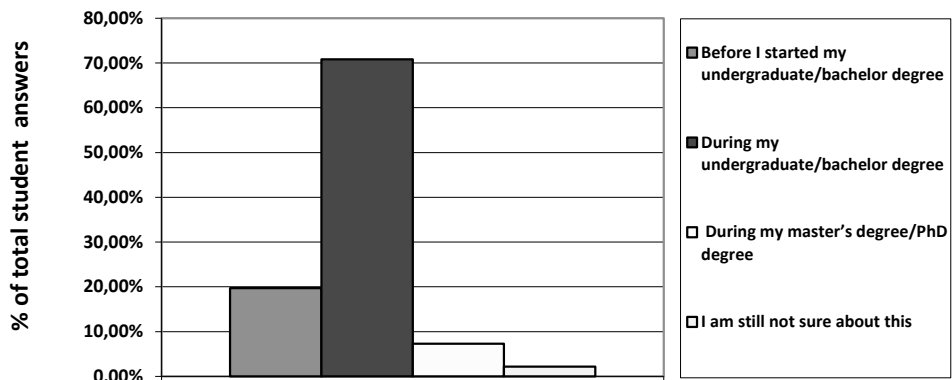


Figure 11.2. Learning to cite and reference

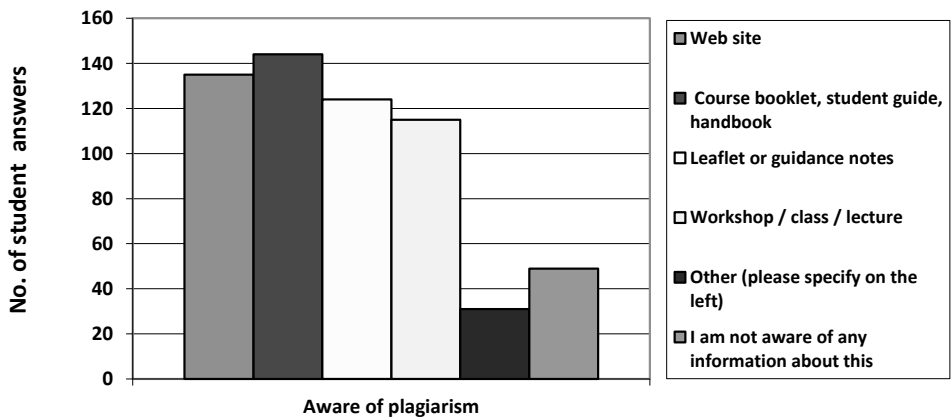


Figure 11.2. Sources of Plagiarism Awareness

These findings suggest that having an institutional formal service to provide students with information and advice on how to avoid plagiarism may be an effective means of raising student awareness and preventing plagiarism. As indicated in Fig. 5, 78.3% of students knew that they would avoid being accused for plagiarism, if they were to use the correct referencing and citation. This is in agreement with what the majority of students consider plagiarism and it would suggest that indeed teaching students about referencing/citing could be the right approach to reduce plagiarism. Of, course, from the results shown in Fig. 6, it does appear that the majority of students (72%) use referencing and citation for promoting their own writing rather than for giving credit to the author(s) of the sourced material. These results confirm the findings above that the students’ definition of plagiarism differs from that of teachers. They also indicate that the students lack of understanding of the importance of acknowledging the authors

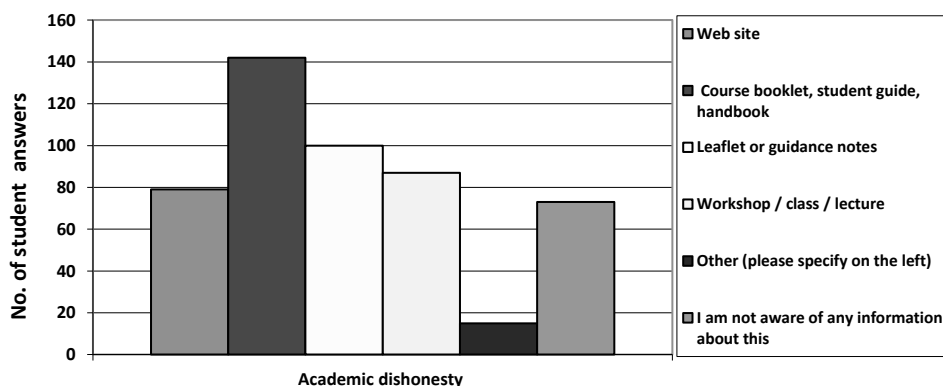


Figure 11.2. Sources for academic dishonesty

Table 3

Responses to type of advising services provided on avoiding plagiarism

Services	Responses
Academic support unit	64
Advice in class during course/module	181
Additional lectures or workshops	80
Advice from tutors or lecturers	222
Guidance from the Library	99
University publisher	91

and crediting them with the originality of the work, can be remedied through teaching the students the correct reasons for using referencing and citations.

Looking now at the reasons for which students plagiarize, the most popular answer (65%) (Fig. 6) was “it is easy to copy and paste form the internet” followed by “they run out of time” (58.5%) and “they think they will not get caught” (58.0%). These results suggest that although students realize that plagiarism is wrong, and know how to cite and reference correctly, they resort to plagiarism simply because it is easy to do if they run out of time, and especially so if they are not going to get caught. Preliminary studies on similar feedback from students who were introduced to the Turnitin software have indicated that they are less likely to plagiarize if they are going to get caught suggesting that the knowledge that plagiarism can be detected can be very effective in preventing plagiarism.

Table 4 presents the findings with respect to the existence of policies, procedures, and penalties for plagiarism and academic dishonesty in the HEIs of Cyprus and with respect to communicating and implementing these.

As Table 4 indicates, the majority of students agree on the existence of policies and procedures for dealing with plagiarism (68.2%) and for dealing with academic integrity (55.4%) at their institution. However, only 32.4% of the students agreed that

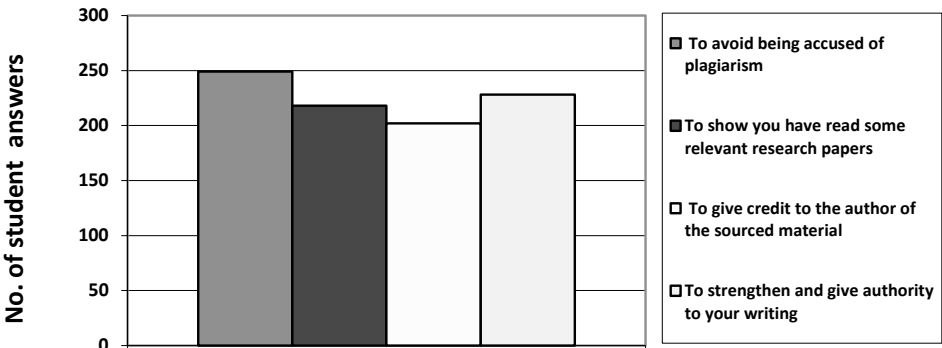


Figure 11.2. Reasons for using correct referencing and citation in scholarly academic writing

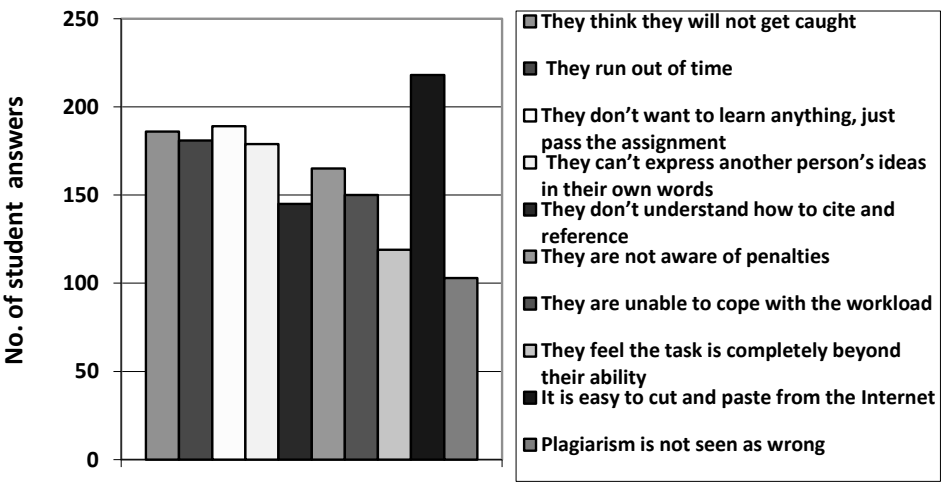


Figure 11.2. What leads students to decide to plagiarize?

policies, procedures and information about penalties for plagiarism are available to them. Furthermore, 43% of the students were not sure how these policies apply. In addition, 41.6% of the students were not sure if all lecturers follow the same procedures for similar cases of plagiarism, or (31.6%) if lecturers treat in the same way cases of plagiarism which are similar, or even (39.5%) whether lecturers follow existing procedures when dealing with plagiarism. These findings suggest (a) that HEIs in Cyprus may not communicate clearly such important information to students, (b) that there may be inconsistency in the ways lecturers apply or follow such policies and procedures, and (c) that there are HEIs that may not always apply their plagiarism related procedures. Noteworthy are also the findings relating to what the penalties for student would be if found guilty of plagiarism, which are shown in Table 5. The most popular (51%) penalty proposed by students for plagiarising in assignments,

Table 4

Percentage of students' answers on Policies, Procedures and Penalties

Question	Strongly Disagree	Disagree	Not sure	Agree	Strongly Agree	Not applicable
The institution where I now study has policies and procedures for dealing with academic dishonesty	1.3%	5.7%	35.8%	37.0%	18.4%	1.9%
The institution where I now study has policies and procedures for dealing with plagiarism	1.0%	4.5%	23.2%	39.2%	29.0%	3.2%
Plagiarism policies, procedures and penalties are available to students	3.2%	12.7%	22.5%	32.4%	22.5%	6.7%
I know what penalties are applied to students for different forms of plagiarism and academic dishonesty	3.8%	19.7%	43.0%	19.4%	5.4%	8.6%
I believe that all teachers follow the same procedures for similar cases of plagiarism	4.85%	11.7%	41.6%	27.6%	11.4%	2.9%
I believe that the way teachers treat plagiarism does not vary from student to student	5.7%	20.6%	31.6%	24.7%	12.7%	4.7%
I believe that when dealing with plagiarism teachers follow the existing procedures	2.5%	14.0%	39.5%	28.3%	12.7%	2.9%

Table 5

Responses of students to the outcome of being found guilty of plagiarism in their assignment or final project/dissertation?

Penalty	Assignment	Final project
No action would be taken	52	29
Verbal warning	153	67
Formal warning letter	116	106
Request to re write it properly	133	159
Zero mark for the work	163	103
Repeat the module or subject	92	101

was “zero mark for the exam”, followed (48%) by “verbal warnings”, whereas for a final project/thesis the most popular (50%) penalty was “request to re-write the work” followed by (33.3%) “formal warning letter”. These findings indicate that students are willing to accept a more severe penalty for the types of student assessment that have less of an impact in their overall grade (i.e. assignment vs. thesis).

Research on the digital tools or other techniques used for detecting plagiarism at the HEIs of Cyprus, showed at that time, that there was no student awareness of any such tools and techniques for detecting plagiarism. This could explain why “not getting caught” is the most popular reason for Cypriot students to plagiarize. The lack of knowledge on the availability and effectiveness of such tools may also explain why they were last in the choices selected by Cypriot students as a means to reduce plagiarism.

Conclusions/Recommendations

According to our findings there is a lack of a clear understanding among Cypriot university students of what plagiarism is. Cypriot universities need to bring about immediate improvements in plagiarism awareness so as to limit plagiarism and address inefficiencies in plagiarism policies, procedures and penalties. Specific recommendations that arise from the above conclusions to the benefit of the Universities and students, are (a) installing into students the concepts of plagiarism, academic dishonesty and the giving of credit to the rightful owner of the original work; (b) implementing in a consistent manner through faculty and audits the institutions' policy and procedures for the prevention and punishment of plagiarism and (c) utilizing plagiarism detection software to deter plagiarism and to enable students to practice on academic writing without plagiarizing. The HEIs in Cyprus have been established only recently (last 25 years) and are by majority under private rather than public governance. It would be of interest to see how the Cyprus results compare with those of Greece who have a much longer public university traditions as well as with those from other EU countries.

References

- Barrett, R., Malcolm, J. (2006) Embedding plagiarism education in the assessment process. *International Journal for Educational Integrity*, 2 (2), 38–45
- Bennett, R. (2005) Factors associated with student plagiarism in a post-1992 university. *Assessment and Evaluation in Higher Education*, 30 (2), 137–162.
- Carroll, J. (2007) *A handbook for deterring plagiarism in higher education* 2nd edition. Oxford: Oxford Brookes University, Oxford Centre for Staff and Learning Development.
- Carroll, J. & Ryan, J. (2005) *Teaching international students: Improving learning for all*. London: Routledge.
- Carroll, J. & Zetterling, C. (2009) *Guiding students away from plagiarism*. KTH Learning Lab.
- Chanock, K. (2008) When students reference plagiarised materials—what can we learn (and what can we do) about their understanding of attribution? *International Journal of Educational Integrity*, 4 (1), 3–16.
- Davis, M. (2007) The role of Turnitin in the formative process of academic writing: A tool for learning and unlearning? *Brookes e-Journal of Learning and Teaching*, 2 (1). Available from: http://bejlt.brookes.ac.uk/article/the_role_of_turnitin_within_the_formative_process_of_academic_writing/ [Accessed 24 April 2013]
- Davis, M. (2009) The role of Turnitin within the formative process of EAP: a tool for global academic culture. BALEAP 2009 Conference Proceedings.
- Emerson, L., Rees, M. & MacKay, B. (2005) Scaffolding academic integrity: Creating a learning context for teaching referencing skills. *Journal of University Learning and Teaching Practice*, 2 (3a), 12–24.

- Hayes, N. & Introna, LD. (2005) Cultural Values, Plagiarism and Fairness: When plagiarism gets in the way of learning. *Journal of Ethics and Behaviour*, 15 (3), 213–23.
- Ireland, C. & English, J. (2011) Let them plagiarise: Developing Academic Writing in a Safe Environment. *Journal of Academic Writing*, 1 (1), 165–172.
- Murray, W. (2006) The plagiarism phenomenon. *E.learning Age*, 22–24. Available from: ABI/INFORM Global [Accessed 24 April 2013]
- McCabe, DL. (2005) Cheating amongst college students: A North American Perspective. *International Journal for Academic Integrity*, 1 (1). Available from: <http://ojs.m1.unisa.edu.au/index.php/IJAEI/article/viewFile/14/9> [Accessed 24 April 2013]
- McCabe, DL., Fenghali, T. & Abdallah, H. (2008) Academic Dishonesty in the Middle East: Individual and contextual Factors. *Research in Higher Education*, 5, 451–467.
- Macdonald, R. & Carroll, J. (2006) Plagiarism: A Complex Issue Requiring a Holistic Institutional Approach. *Assessment and Evaluation in Higher Education*, 31 (2), 233–245.
- Neville, C. (2010) *The Complete Guide to Referencing and Avoiding Plagiarism*. Open University Press, Berkshire: McGraw Hill.

Acknowledgements

Research in this paper has been funded by the project Impact of Policies for Plagiarism in Higher Education across Europe IPPHEAE 510321-LLP-1-2010-1-UK-ERASMUS-EMHE. Authors are very grateful to all the project consortium partners for their cooperation in survey design and data collection: Coventry University (UK), Technical University of Łódź (Poland), Aleksandras Stulginkis University (Lithuania), Mendel University in Brno, Czech Republic.

Authors

Angelika Kokkinaki, kokkinaki.a@unic.ac.cy,
Melpo Iacovidou, iacovidou.m@unic.ac.cy,
Catherine Demoliou, demoliou.c@unic.ac.cy,
University of Nicosia, Cyprus

Copyright © 2013 Authors listed on page 192: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

STRATEGIES AND RESPONSES TO PLAGIARISM IN SLOVAKIA

Július Kravjar, Juraj Noge

Abstract: The rapid increase in the number of higher education institutions (HEI), students, ICT and internet penetration after 1989 and a low level of copyright and intellectual property rights awareness contributed to the growth of plagiarism at HEIs in Slovakia. In 2008, the Ministry of Education decided to establish a common national repository for bachelor, master, dissertation, rigorous and habilitation theses. The amendment to the HEI Act dated 2009 stipulates that every thesis or dissertation of a HEI operating under the Slovak legal order shall be sent to the central repository and undergo the originality check before it is defended. The originality check of every new thesis is made against the national repository, as well as against other resources, primarily the internet ones. From 2010 to the end of March 2013, more than 235 thousand theses and dissertations have gone through this process. The quality of the solution (organisational and technical) meets the expectations and has a positive impact on the academic community. This paper describes the implementation of the Central Repository of Theses and Dissertations and of the Plagiarism Detection System into practice and the experience from their operation. It also analyses the impacts of the implementation on HEIs, feedback and reactions of the academia and the public, and provides recommendations regarding the prevention and detection of plagiarism. It also addresses the growth in the number of internet texts and texts in media containing the term *plagiátorstvo* (plagiarism) and the growth of search hits (search frequency) for this term in the period 2002–2012.

Introduction

The paper analyses the facts that contributed to the spread of plagiarism at HEIs in Slovakia and the responses to the spreading of plagiarism. The academic community and Ministry of Education (MinEdu) realised the need to stop the thriving plagiarism. Time was needed for the awareness to grow into the belief that the fight against plagiarism requires systemic action on a national level. Finally, nationwide central repository of theses and dissertations and nationwide anti-plagiarism system were implemented.

In the second part of the paper, in relation to the implementation of both systems mentioned above, we analyse the number of texts containing the word *plagiátorstvo* (*plagiarism*) on the internet and in the media for the period 2002–2012 and the behaviour of the community regarding the frequency of the searches for the word *plagiátorstvo* (*plagiarism*). The aim was to identify whether the adoption of the systemic measure at the national level in 2010 maximised the number of texts and their increases for the period up to 2010 inclusive. A similar question was raised regarding the frequency of searches for the word *plagiátorstvo* (*plagiarism*) in the examined period.¹

¹The paper is the English analogy of the paper of Kravjar (2013b)

Abbreviations

CR:	Central Repository of Bachelor's, Master's, Dissertation, Rigorous and Habilitation Theses
APS:	Anti-plagiarism System (Plagiarism Detection System, System for Originality Check)
HEI:	Higher Education Institution
SRC:	Slovak Rector's Conference
MinEdu:	Ministry of Education, Science, Research and Sport of the Slovak Republic
SCSTI:	Slovak Centre of Scientific and Technical Information (CVTI SR: Centrum vedecko-technických informácií Slovenskej republiky)

Explanation

A “small doctorate” can be received by a person with a master's degree. It requires that a candidate passes rigorous examination and defends his/her rigorous thesis.

A prerequisite for the granting of the scientific-pedagogical degree “docent” (assistant professor) is habilitation lecture, submission and defence of the habilitation thesis.

Background Information

Low awareness of copyright, intellectual property rights and academic ethics combined with a growing number of internet users, students and higher education institutions has been a breeding ground for the spread of plagiarism in written papers at HEIs. According to Skalka et al. (2009b)

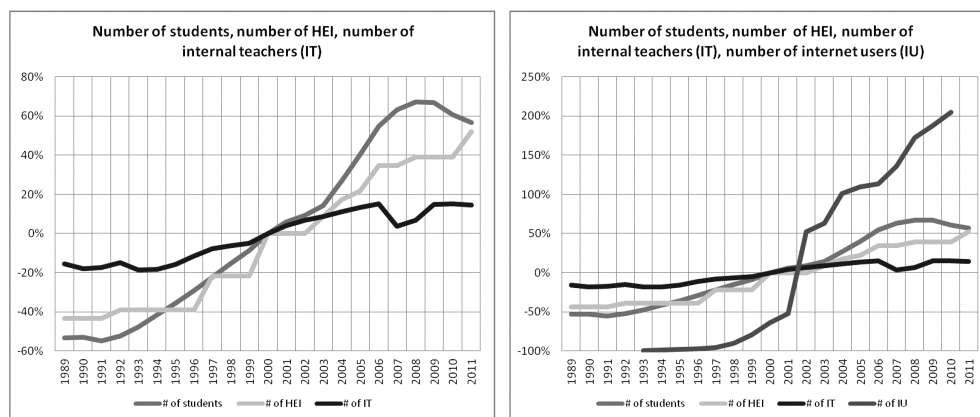
Internet was the likely catalyst for an avalanche of plagiarism; it provided students, often including primary school pupils, with resources for quick, easy and effortless access to information about the assigned topic and later even complete texts.

A comprehensive overview of the state of academic ethics in Slovakia is analysed by R. Králiková (2009) in her publication “Implementation of rules of academic ethics at Slovak HEIs”. She states that:

The issue of academic ethics started to attract attention in Slovakia especially in the past three years. It usually appeared in relation to the issue of plagiarism among students and teachers. However, plagiarism is just one of the infractions against academic ethics and the issue of academic ethics in Slovakia is much broader than just plagiarism.

K. Staroová (2010) describes the causes supporting plagiarism as follows:

... plagiarism has its roots mainly in the method of teaching; the absence of systematic education (ideally from the primary school level), teaching the students to write argumentative texts, list resources correctly, benefits of quotations (it is much easier to go back to the original source and critically consider the author's conclusions) and especially ethical principles that would clearly define plagiarism as violation of intellectual property rights and parasitizing on the work of others and not playing down the situation. It is very sad to watch how many teachers who should teach students



Graph 1 and 2. Development of observed indicators over time transformed according to the formula $100 \cdot (x_i - x_{\text{median}}) / x_{\text{median}}$

ethical principles, explain to them the nature of plagiarism and require them to think and write originally are themselves guilty of the sin and play down its significance if caught in the act.

In 1989, there were 13 HEIs in Slovakia and today there are 40; 36 operate under the Slovak legal order. The development of the number of students, teachers, HEIs (iEDU, 2012) and the number of internet users (www.indexmundi.com, 2012) is shown in Graph 1 and 2. The number of students and teachers apply to HEIs operating under the Slovak legal order.

The year 2000 was the year of highest annual increase in the number of HEIs (5), 2002 was the year with the highest increase in the number of internet users (nearly 1.5 million, almost 28% of the total population). In 2006, there was the highest annual increase in the number of students (over 20,000); 2004 and 2005 were also the years of high annual increases in the number of students (more than 17,000 and more than 18,000 respectively). The highest increase in the number of teachers (774) was recorded in 2009.

In Slovakia, there were many discussions on the subject of plagiarism, but without a specific result for a long time. In 2001, the first HEI (Vysoká škola manažmentu—in Trenčín) in Slovakia started to use an electronic system to detect plagiarism. It was not only the system to detect plagiarism itself; there were also other related measures: the rules of academic ethics were adopted, as well as a process defining the procedure for the investigation of suspected plagiarism (Kročitý, Argaláš, 2012).

In 2006, more than 213,000 students studied at Slovak HEIs for—3.3 times the number in 1989, there were 32 HEIs—2.4 times the number in 1989; the number of internet users increased and exceeded 3 million, representing more than 55% of the total population. However, still only one HEI used an electronic system to detect plagiarism. In 2008, it was joined by another HEI (University of Economics in Bratislava) and in 2009 by a third HEI (Comenius University in Bratislava).

Opinions of representatives of HEIs on the use of the system to detect plagiarism varied initially. A significant progress in this area was SRC plenary's opinion on plagiarism dated 28 September 2006 entitled "Measures to eliminate plagiarism in the processing and presentation of bachelor's, master's and dissertation theses" (Slovak Rectors' Conference, 2006a). The document is important in that:

- Plagiarism is considered a serious problem that must be addressed by informing students how to write papers correctly and how to cite literary sources;
- The best way to reduce plagiarism is prevention;
- In the case of confirmed plagiarism, it requires that the student bear the consequences in accordance with the internal regulations of a HEI.

Another important document approved by SRC on the same day, "The Code of Ethics for HEI Employees" (Slovak Rectors' Conference, 2006b). However, the Code contains no mention of plagiarism, as if plagiarism never related to teaching and research staff of HEIs.

These two documents did not change the status quo nor did they affect the suppression of plagiarism. A more effective action was needed, with a major impact on combating plagiarism. SRC (2008) adopted a resolution with a potential of a fundamental solution:

The plenary of Slovak Rectors' Conference revisited the issue of plagiarism. It asked the Ministry of Education of the Slovak Republic to coordinate the relevant activities, especially those relating to the acquisition of the software. It also urges the members of the Slovak Rectors' Conference to create an electronic archive of theses. It recommends that HEIs modify their regulations to address the issue of plagiarism.

In March 2008, internet media reported (Supuka, 2008) that "the Minister of Education promised to obtain software to control plagiarism for HEIs in Slovakia". The Ministry of Education decided that a comprehensive solution will be implemented at a national level covering the collection, processing and originality check of specified papers. HEIs in Slovakia will be required to use this comprehensive solution (Skalka et al., 2009a).

The basic strategic goal was defined and it was necessary to develop a strategy to achieve it. One of the major factors that contributed to the success in achieving the strategic objective was finding the support in the Parliament and incorporation in the amendment to the Higher Education Act of the obligations for HEIs to send specified papers in the central repository in order to be checked for originality after registration.

The amendment to the Higher Education Act was approved in 2009 and paved the way for the preparation of a nationwide repository of bachelor, master, dissertation, rigorous and habilitation theses. SCSTI was commissioned by the Ministry of Education to obtain the necessary software by public procurement and it was also made responsible for its operation. The preparation phase was challenging due to a lack of time. The real operation of the central repository and system to detect plagiarism began on 30 April 2010.

Important Moments

The important moments associated with building and operating of both systems are shown below (more details can be found in the paper “Barrier to thriving plagiarism” (Kravjar, 2012)):

- Absence of systemic barrier for plagiarism fight at HEIs;
- Agreement of rectors of HEIs regarding the system for the detection of plagiarism;
- Support for the rectors’ requirement to obtain a system to detect plagiarism from MinEdu;
- Elaboration of the study “Prevencia a odhaľovanie plagiátorstva”² (Skalka, J. et al., 2009c);
- Legislative support from the Parliament—approval of an amendment to the Higher Education Act;
- Appointment of SCSTI as the operator of both systems
- Preparation of methodological materials for HEIs;
- Preparation of the project;
- Procurement of the system, technical, technological and organisational preparation of the systems;
- Effective cooperation in the preparation phase between MinEdu, SCSTI and the supplier of the system SVOP, s. r. o.;
- On 30 April 2010, the routine operation of the CR and APS was started at SCSTI’s premises; all HEIs operating in Slovakia under Slovak legislation are required to send, before the defence, bachelor’s, master’s, dissertation, rigorous and habilitation theses to the CR, where it is archived for 70 years; each work is checked for originality and originality check protocol is delivered to HEIs electronically;
- The new algorithm for the originality check won first place at international competition External Plagiarism Detection Performance at PAN 2011 Lab Uncovering Plagiarism, Authorship, and Social Software Misuse Conference in Amsterdam; it was not only the overall victory, but total victory in all four evaluation indicators;
- Second place at the International Congress ITAPA 2011 (Information Technology and Public Administration) in the category New services;
- Migration of both systems in 2011 to the Data Centre for Research and Development with more efficient, more reliable and more secure ICT infrastructure; the Data Centre is built by SCSTI within the national project “Infrastructure for Research and Development—Data Centre for Research and Development (CVTI SR, 2010)”;
- A change in the system was introduced by another amendment to the Higher Education Act, which imposed on the operator an obligation to make papers stored in the CR available to the public; papers registered in the CR after 31 August 2011 are made available;

²“Prevention and detection of plagiarism”

- In February 2013, we signed up for the competition “European Prize for Innovation in Public Administration” (in the category “Support to Science and Education”) which was announced by the European Commission. In the end of March, we were informed that we had proceeded further and on April 11th we received the information that our application had been shortlisted as one of the 6 most innovative initiatives in our category. The jury will make the final selection on the winning three initiatives in our category based on a hearing and interview.

Prepared Changes

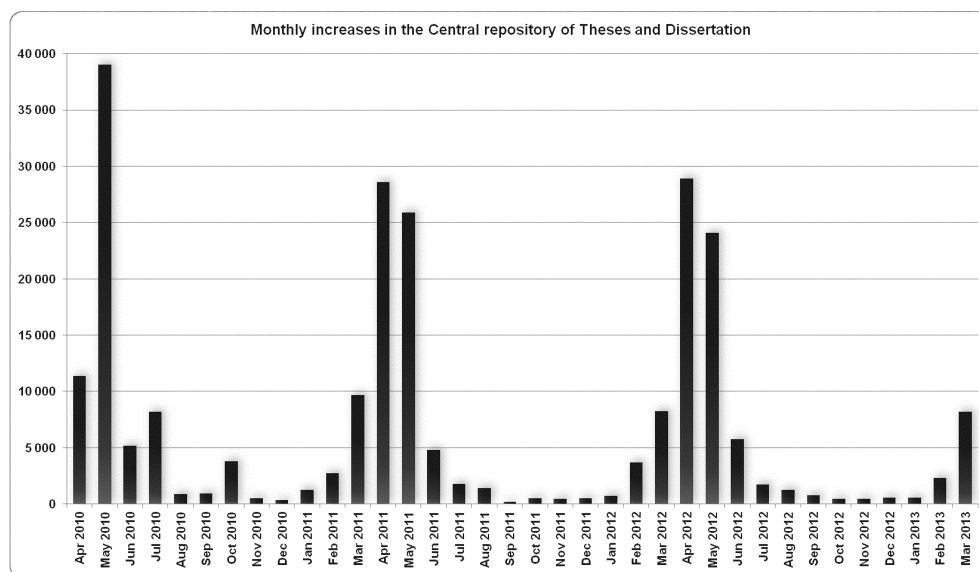
CR and APS are live and developing systems responding to everyday operation impulses, as well as to technological trends in the area of plagiarism detection software. We are prepared to develop ANTIPLAG further. Only the most important plans are mentioned:

- Deployment of the winning originality check algorithm from the international competition in Amsterdam (www.uni-weimar.de, 2011). This competition involved only English, Spanish and German texts. The algorithm won overall, as well as in each of the four evaluated indicators. One of the indicators was detection of translational plagiarism (cross-languages plagiarism). According to professor Weber-Wulff (Pataki, 2012), who has dealt with plagiarism since 2002, detection of translational plagiarism belongs to greatest weaknesses of anti-plagiarism systems. This claim is based on her experience with testing 48 of such systems from different countries. Detection of translational plagiarism will make the new version of ANTIPLAG a top system in its class.
- The originality check protocol in its paper version will have a more advanced “brother”: a web version with all text matches that ANTIPLAG detected, not only text matches from the first fifteen documents with the highest percentage of over-the-threshold match (which limits the paper version of the protocol).
- Extension of the comparison corpus with papers in the Czech language from internet sources.
- Preparation to extend the comparison corpus with papers in the English language from internet sources.

ANTIPLAG is used by HEIs 24/7 throughout the year. The seasonality of additions in the number of papers in the CRTD is illustrated in Graph 3.

Reactions to the Implementation of CR and APS into Routine Operation

According to the statements of HEI officials in the media, the mere publication of the information that operation of such systems is being prepared had a positive preventive effect. Students and teachers realised that their papers will be under permanent supervision of the plagiarism detection system and they began to approach the creation of papers and work with literature more seriously.



Graph 3. Annual cycles of theses and dissertation increases in the CR

Reactions to the implementation of the CR and APS systems in the media agree that they contributed to:

- Increased responsibility of students and teachers;
- Greater student autonomy in the creation of papers;
- Improvement in the quality of papers;
- A higher level of citations.

One quote for all (www.topky.sk, 2010):

Zuzana Pucherová from the Department of Ecology and Environmental Sciences at the Constantine the Philosopher University in Nitra noted that the anti-plagiarism system forced students and teachers to be more responsible. “A teacher has to approach the reading of partial texts with greater precision in order to avoid overlooking something,” she commented. According to her, although a student is the author, the responsibility is borne by teachers/tutors.

The implementation of both systems was a significant step; however, the originality check only covers bachelor’s, master’s, dissertation, rigorous and habilitation theses. There still remains a range of publications produced by HELs, which are not covered by the originality check. A critique of this situation and plagiarism among teachers was included in E. Mistrik’s (2012) article “Why academics remain silent about plagiarism”. He looks for and finds reasons why academics remain silent. He reveals why some teachers resort to unethical practices and specifies them. He also writes:

However, I feel that the most widespread form of plagiarism is using the papers of their students/postgraduates (“But I taught him”, “But he/she did under my

supervision”), or of their subordinates (“After all, he/she did in my laboratory”, “But he/she is in my department”).

It is a pity that the amendments to the Higher Education Act did not consider feedback. HEIs are not required to disclose information, such as the number of suspected plagiarism cases investigated, how many of them represented a positive finding, and what were the sanctions against the plagiarists. Even with a nationwide system in place, we do not know whether the number of plagiarists is increasing or decreasing. There is a lack of exact data on plagiarism.

The Occurrence of the Term *plagiátorstvo* (*plagiarism*) in Texts on the Internet and in the Media

The launch of CR along with APS into live operation in 2010 represents a significant milestone in the fight against plagiarism of theses (bachelor, master, rigorous, dissertation, habilitation) at HEIs in Slovakia. The pre-implementation period and the first months of the operation of these systems was a period when *plagiarism* often appeared in the traditional media and on the internet.

Our objective was to determine whether this breakthrough event affected the number of texts containing the word *plagiátorstvo* (*plagiarism*), as well as their annual increase on the internet and in the media in the period 2002–2010. We assumed 2010 to be the year with both indicators at the maximum.

The base year 2002 was selected based on a recommendation (Ionescu, 2010): Search for a specified time interval: “This option works best for pages published after 2001”. We used Google advanced search with the following parameters:

- Search in Slovak language;
- Search in Slovakia;
- Search only in the text of pages (not in the page title, not in the URL address, not in the links to the page);
- Strictly safe search;
- Last update at any time;
- Any format;
- Rights to use unfiltered by license;
- Unspecified site or domain.

Given that the search for individual years provided either “*Number of results*” or “*Approximate number of results*”, we performed the search on a monthly basis in order to reach homogeneity—the search engine always gave “Number of results”. After displaying “Number of results”, we always checked if there is no message after the last result saying: “*In order to show you the most relevant results, we have omitted some entries very similar to the N already displayed. If you like, you can **repeat the search with the omitted results included.***” If there was such a message, the search was repeated and “Number of results” afterwards was considered final.

The 2002–2006 period may be characterised as a period of very low occurrence of information about plagiarism on the internet with given search parameters. This

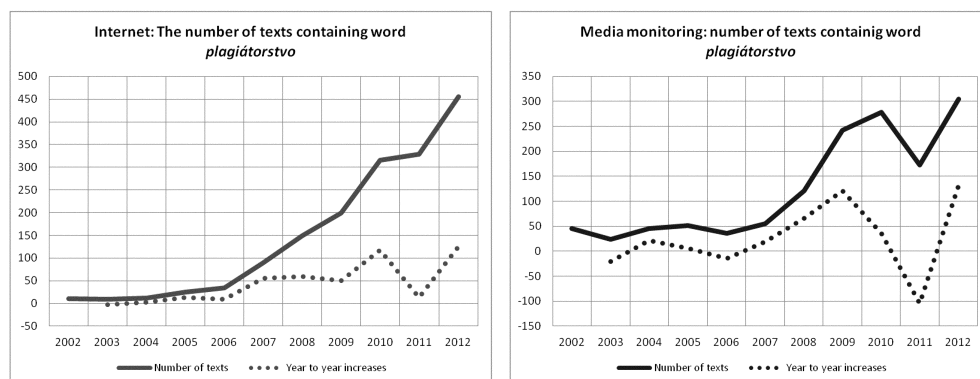


Figure 11.2. Graph 4. Number of texts on the Internet; Graph 5. Number of texts in the media

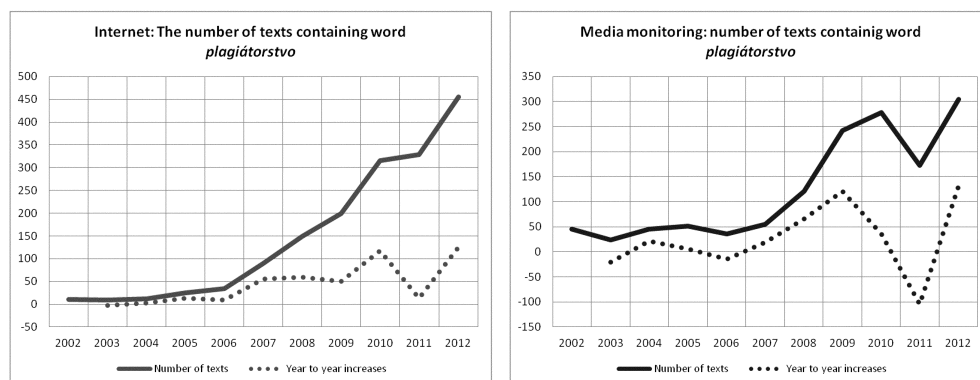


Figure 11.2. Graf 6. Media monitoring: Numbers; Graf 7. Media monitoring: Increases

corresponds with the finding of R. Králiková (2009). However, the internet penetration amounted to 40% and 56% in 2002 and 2006 respectively. A sharp increase in the number of internet texts containing the word *plagiátorstvo* (plagiarism) occurred after 2006 (Graphs 4 and 5).

The data relating to the media monitoring was provided by Storin (Bartko, 2012, 2013). The occurrence of the variants of the word *plagiátorstvo* in the media was monitored (radio, television, press, internet media). The monitoring of radio and press covers the 2002–2012 period, the television monitoring started in 2004 and internet media monitoring started in 2007. Like with the internet search, a very slow growth in the number of texts on plagiarism is characteristic for the initial years 2002–2007 (Graphs 6 and 7).

We assumed that the frequency of texts with the word *plagiátorstvo* on the internet, in media monitoring and the frequency of searches for this word by internet users for the period until 2010 (including their increases) will culminate in 2010. This assumption

was confirmed only partially. The results overview is shown in Table 1; the columns “stagnation” and “moderate growth” do not apply to the last three columns of the table.

Regarding the texts containing the word *plagiátorstvo* on the internet and in the media, only in one case for the 2002–2010 period, the maximum is not achieved in 2010—in the case of press monitoring (maximum in 2009). Regarding the increases, there are four exceptions: overall monitoring, radio and press (maximum in 2009), television monitoring (maximum in 2008). The number of internet texts containing the word *plagiátorstvo* shows a rising trend after 2010, contrary to the media.

The maximum of internet texts with the word *plagiátorstvo* was achieved in December 2010, the value over 50% of the maximum was achieved in November 2009. Values over 40% of the maximum were achieved in October, November 2009 and in March, October, November and December 2010. The maximum increase in the number of texts was not concentrated in the first month of using CRTD and APS, but in the period seven months later.

The Number of Search Hits for the Word *plagiátorstvo* on the Internet

On one hand, interest in plagiarism leads to an increasing number of texts about plagiarism on the internet; on the other hand, there is the activity of internet users searching for the word *plagiátorstvo* (*plagiarism*). Google Trends (www.google.com/trends/) provides the ability to quantify the frequency of searched words entered for a period since 2004 per months in percentage compared to the month in which the maximum number of searches was recorded. It should be noted that the zero value does not mean any query, but that the number did not exceed the threshold value. We assumed that the interest in the search for the word *plagiátorstvo* would peak in 2010.

When entering the word *plagiátorstvo* in Google Trends and defining Slovakia, we got a graph with only one over-the-threshold value (June 2011). Given that internet users largely ignore diacritics when entering search queries, we also looked at the trends in searching for the word *plagiatorstvo*. The maximum number of queries (Graph 8) was recorded in May 2010; over-the-threshold values are shown in January to April 2010. An increased interest for the search grows every year in the period of submission of papers.

Google Trends allows identifying not only relative numbers of searches for one word, but also for multiple words. When identifying the number of queries for the words *plagiatorstvo* and *plagiátorstvo* (Graph 9), we get a higher number of over-the-threshold values (a sum of two below-the-threshold values exceeding the threshold value). Results for *plagiátorstvo* and *plagiatorstvo* (Graph 10) are not identical with the results for *plagiatorstvo* and *plagiátorstvo*; however, the maximums are identified in the same months.

Despite the “non-commutativity” in identifying the number of searches, the trends are well-identified. Identifying the number of searches are not invariant in time; a shift in identification by 24 hours identifies small deviations compared with the previous day.

Results regarding the number of searches for the word *plagiátorstvo* document that the use of Slovak language without diacritics when entering search queries dominates the grammatically correct Slovak language.

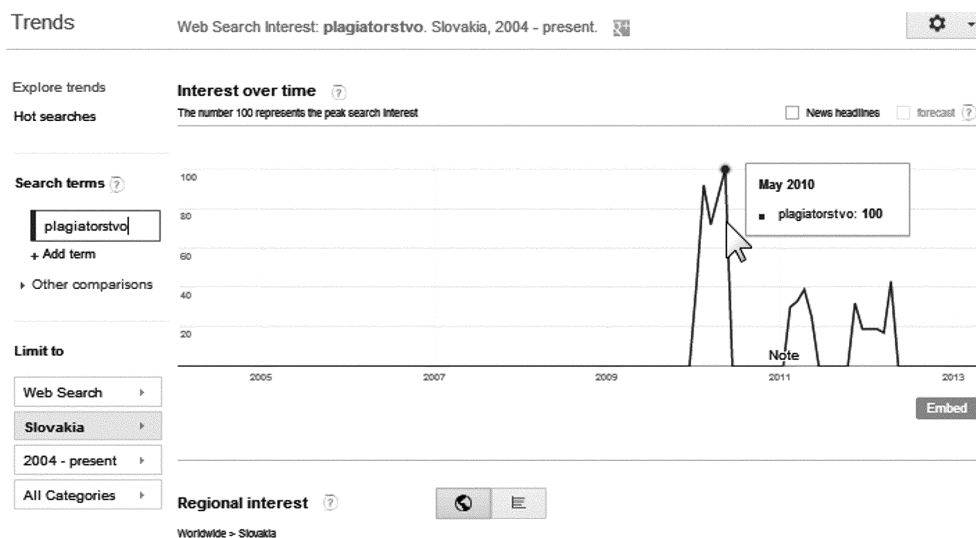


Figure 11.2. Graph 8. Relative number of searches for the word *plagiatorstvo*

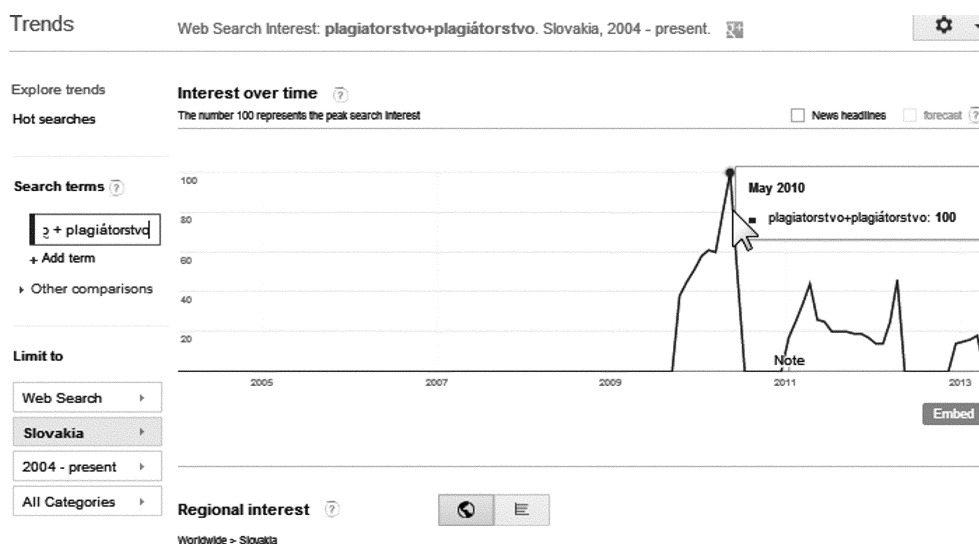


Figure 11.2. Graph 9 Relative number of searches for the word *plagiatorstvo* and *plagiátorstvo*

The first month of the use of CRTD and APS was May 2010 and May showed the maximum number of searches. The local maximums of the number of searches have a year's periodicity and they belong to the main period of submission of papers (March to May).

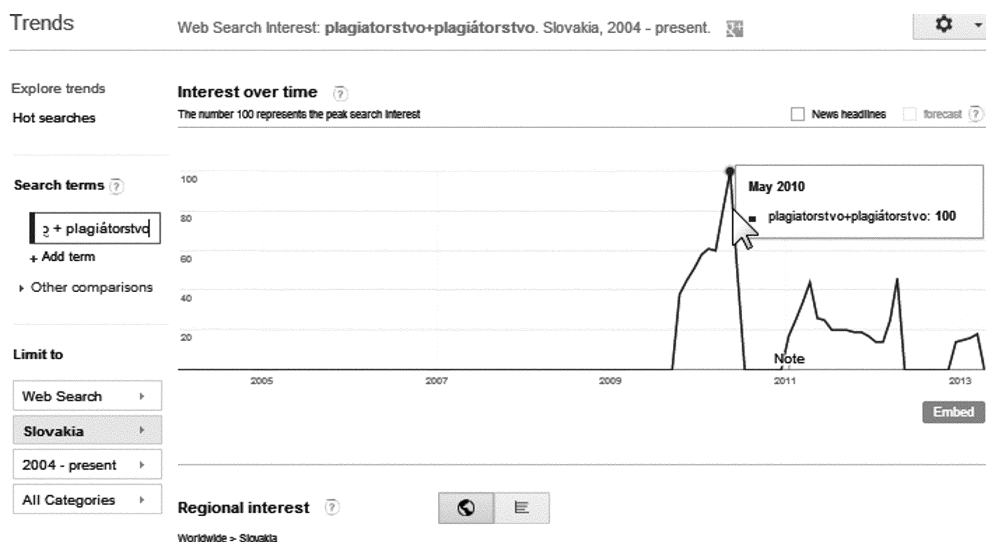


Figure 11.2. Graph 10. Relative number of searches for the word *plagiátorstvo* and *plagiatorstvo*

Table 1 summarises results regarding the maximums of texts containing the word *plagiátorstvo* on the internet, in the media and related to the maximum interest in searching for the word *plagiátorstvo* or *plagiatorstvo*.

Summary

The implementation of the strategic objective to deploy national CR and APS was a fundamental systemic step and brought positive effects in the fight against the spreading of plagiarism. The operation of both systems is a preventive tool; it contributes to improvements in education, to a better understanding and application of academic ethics, copyright and intellectual property rights, and to improvements in students' work with literature. The papers registered in CRTD from 1 September 2011 are available to the public (www.crzp.sk/crzpopacxe). The public is able to verify suspected plagiarism. We can conclude that theses and dissertation are under three types of control: the tutor's control, the examination committee's control and the public control.

However, we do not have exact plagiarism data due to non-existent feedback from HEIs regarding the results of investigations into suspected plagiarism.

The originality check does not cover all HEI publications; there is still room to implement the originality check for other types of HEI publications.

The use of plagiarism detection systems has had longer tradition in many countries than in Slovakia. The implementation of CR and APS and their mandatory use nationwide (at HEIs operating under Slovak legislation) is a unique solution in Europe and very likely in the world.

Table 1

A brief overview of the results on the internet, media monitoring and the number of searches

The results overview	Number of texts				Relative search number
	Stagnation	Moderate growth	Maximum		
			Number	Increases	Number
Texts on the Internet —annual base	2002–2004	2005–2006	2010	2010	
Texts on the Internet —semiannual base	2002–2004	2005–2006	2010 2nd half-year	2010 2nd half-year	
Texts on the Internet —quartal base	2002–2004	2005–2006	2010 4th quartal	2010 4th quartal	
Texts on the Internet —monthly base	2002–2004	2005–2006	2010 December	2010 December	2010 May
Media monitoring total —year base	2004–2007	—	2010	2009	
Media monitoring: radio —annual base	2004–2007	2008–2010	2010	2009	
Media monitoring: TV —annual base	2004–2007	2008–2010	2010	2008	
Media monitoring: press —annual base	2002–2007	—	2009	2009	
Media monitoring: i-media —annual base	—	2007–2009	2010	2010	

Education and prevention play an important role in combating plagiarism. HEIs are aware of this and improve the moral and knowledge level of students in this regard. However, there is one major drawback: our Higher Education Act does not recognise a withdrawal of a university degree. If moral principles are violated in the preparation of the thesis and it remains undetected until the award of a degree, the degree remains intact in the hands of the owner (Húska, 2012). Therefore, we have seen no withdrawal of degrees yet, for example, in Germany (case Guttenberg, Schavan) and in Hungary (case Schmitt). Hopefully a change is in sight—in August 2012, the newspapers informed (TASR.SK, 2012) that a fraudulently obtained degree will be qualified as a criminal offense. The issue should be subject to the amendment of the Criminal Code.

The number of text containing the word *plagiátorstvo* was very low on the Slovak internet and in the media until 2006 (internet search with the parameters specified above) or until 2007 (media) which corresponded with the number of searches for this word. This was a reflection of the situation of the whole society—the issue of plagiarism, copyright, intellectual property rights was not given sufficient attention by the society. Significant changes started in 2007/2008, when the number of texts on

plagiarism significantly increased, and the issue of plagiarism became a public matter. The term *plagiarism* is no longer a Cinderella of Slovak internet and media. We are more open to the issue of plagiarism, its presence is openly admitted and there is much more written and spoken word on the topic. The interest in the search for the word *plagiátorstvo* or *plagiatorstvo* culminates in spring every year which is the main period for submission of theses and dissertations (Graphs 3, 8, 9 and 10). A large contribution to that was the effort that led to the nationwide implementation of CR and APS.

References

- Bartko, V. (2012). bartko@storin.sk. RE: rešerš. 26.11.2012.
- Bartko, V. (2013). bartko@storin.sk. RE: rešerš. 14.1.2013.
- CVTI SR (2010). *Infraštruktúra pre výskum a vývoj—Dátové centrum pre výskum a vývoj*. [ONLINE] Available at: <http://dc.cvtisr.sk/index.php?&menu=2>. [Accessed 31.1.2013]
- Google (2012). *How is the data derived?*. [online] Available at: <http://support.google.com/trends/answer/92768?hl=en&>. [Accessed 31.1.2013]
- Húska, M. (2012). *Napišem vám diplomovku. Stačí zaplatiť..* [online] Available at: http://style.hnonline.sk/c3-55146090-kt0000_d-napisem-vam-diplomovku-staci-zaplatit. [Accessed 24.1.2013]
- iEDU (2012). *Časové rady*. [online] Available at: <http://www.uips.sk/statistiky/casove-rady>. [Accessed 15.12.2012]
- Ionescu, M. (2010). [online] Available at: <http://www.metrolic.com/about-your-web-history-117749/> [Accessed 22.1.2013]
- Králiková, R. (2009). *Zavádzanie pravidiel akademickej etiky na slovenských vysokých školách*. [online] Available at: <http://www.governance.sk/index.php?id=1193> [Accessed 24.1.2013]
- Kravjar, J. (2012). *Barrier to thriving plagiarism*. Conference paper—The 5th International Plagiarism Conference. 2012. [online] Available at: http://archive.plagiarismadvice.org/documents/conference2012/finalpapers/Kravjar_fullpaper.pdf. [Accessed 14.1.2013]
- Kravjar, J. (2013a). *Výskyt slova plagiátorstvo v textoch na slovenskom internete*. [online] Available at: <https://www.vedatechnika.sk/Blog/Lists/Posts/Post.aspx?ID=86>. [Accessed 21.1.2013]
- Kravjar, J. (2013b). CRZP/APS: míľniky, aktuálny stav, pripravované zmeny. Symposium paper—37th Symposium INFOS 2013
- Kročitý, P. & Argaláš, M. (2010). *10-ročné skúsenosti s bojom proti plagiátorstvu na Vysokej škole manažmentu v Trenčíne*. [online] Available at: <http://www.apvv.sk/buxus/docs/aktuality/konferencia-etika/konferencia-20100414-krocity-argalas.pdf>. [Accessed 21.1.2013]
- Mistrík, E. (2012). *Prečo akademici mlčia o plagiátorstve*. [online] Available at: <http://mistrík.blog.sme.sk/c/299877/Preco-akademici-mlcia-o-plagiatorstve.html>. [Accessed 22.1.2013]
- Skalka, J., Drlík, M., Grman, J. & Vozár, L. (2009a). *Záverečné práce a plagiátorstvo*. [online] Available at: <http://www.fem.uniag.sk/uninfos2009/sites/sk.uninfos2009/files/skalka.pdf>. [Accessed 24.1.2013]
- Skalka, J., Drlík, M. & Vozár, L. (2009b). *Korene plagiátorstva*. [ONLINE] Available at: http://www.fem.uniag.sk/sit2009/zbornik/13_skalka.pdf. [Accessed 24.1.2013]

- Slovenská rektorská konferencia (2006a). *Stanovisko pléna SRK k plagiátorstvu*. [ONLINE] Available at: <http://old.srk.sk/zaznam/90/Stanovisko-plena-SRK-k-plagiatorstvu/>. [Accessed 21.1.2013].
- Slovenská rektorská konferencia (2006b). *Etický kódex zamestnancov vysokých škôl*. [online] Available at: <http://old.srk.sk/zaznam/89/Eticky-kodex-zamestnancov-vysokych-sk%C3%B4l/>. [Accessed 21.1.2013]
- Slovenská rektorská konferencia (2008). *Zápisnica zo 43. riadneho zasadnutia pléna Slovenskej rektorskej konferencie, ktoré sa konalo 8. 2. 2008 v Trenianskych Tepliciach*. [online] Available at: http://old.srk.sk/files/file_25.pdf. [Accessed 21.1.2013]
- SME.SK (2012). *Podvodne získaný titul bude trestným činom*. [ONLINE] Available at: <http://www.sme.sk/c/6511466/podvodne-ziskany-titul-bude-trestnym-cinom.html>. [Accessed 21.3.2013]
- Staroňová, K. (2010). *Plagiátorstvo*. [online] Available at: <http://www.governance.sk/index.php?id=1731>. [Accessed 24.1.2013]
- Supuka, J. (2008). *Akademická etika a plagiátorstvo*. [online] Available at: <http://www.polnohospodar.sk/kategorie-spravodajstva/24-ponohospodar-1252/476-akademicka-etika-a-plagiatorstvo>. [Accessed 21.1.2013]
- www.indexmundi.com (2012). *Internet users*. [online] Available at: <http://www.indexmundi.com/slovakia/internet-users.html>. [Accessed 21.1.2013]
- www.topky.sk (2010). *Antiplagiátorský systém vystrašil študentov, viac citujú*. [online] Available at: <http://www.topky.sk/c1/10/784420/Antiplagiatorsky-system-vystrasil-studentov--viac-cituju>. [Accessed 21.1.2013]
- www.uni-weimar.de (2011). *Evaluation Results*. [online] Available at: <http://www.uni-weimar.de/medien/webis/research/events/pan-11/pan11-web/plagiarism-detection.html#results>. [Accessed 21.3.2013]
- www.zive.sk (2011). *Cena ITAPA 2011: Jedným z najlepších je projekt antiplagiátorského systému*. [online] Available at: <http://www.zive.sk/tlacove-spravy/cena-itapa-2011-jednym-z-najlepsich-je-projekt-antiplagiatorskeho-systemu/sc-5-a-296918/default.aspx>. [Accessed 4.4.2013]

Authors

Július Kravjar, julius.kravjar@cvtisr.sk,
Juraj Noge, nogaj@cvtisr.sk,
Slovak Centre of Scientific and Technical Information, Slovakia

Copyright © 2013 The authors listed on page 201 grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

INTER-UNIVERSITY COOPERATION ON PLAGIARISM DETECTION SYSTEMS IN CZECH REPUBLIC

Luboš Lunter, Daniel Jakubík, Šimon Suchomel, Michal Brandejs

Abstract: Plagiarism detection systems can help to discover plagiarised material by finding similar documents. The field of plagiarism detection has been dealt with by Masaryk University since 2006 when the users of the Information System of Masaryk University (IS MU) started using e-learning tools including electronic study materials. In 2008, the joint initiative of most public universities in the Czech Republic resulted in development of the National Register of Theses and plagiarism-tracking system Theses.cz. This was followed by the start of the system tracing plagiarisms in seminar papers Odevzdej.cz. Repozitar.cz continues the inter-university control of plagiarisms in the field of scientific and professional publications. The paper shows how these systems work and discuss the benefits of inter-university cooperation. Plagiarism analysis of the document can detect the suspicious similarities with identified source document.

Introduction

In the past few years there is an obvious trend of keeping and processing documents electronically. The main source of electronically available documents is the Internet. It contains not only Web pages, but also vast collection of other text documents such as scanned books or archived theses. There are lots of benefits in having documents electronically available and easily accessible, but it also increases the risk of misuses for plagiarism.

The area of plagiarism is very huge and it can cover many creative human activities in lots of different forms. However, it is very easy to plagiarise an electronic document on contrary to detection of corresponding falsification. Therefore it is important to develop and provide some automated systems which can help to detect potentially plagiarized documents. As reaction for this increasing need, several anti-plagiarism systems have been developed (Brandejs, M. et al., 2009). For purposes of this paper, the anti-plagiarism system will be understood as a service focused on discovering potential plagiarism in a given document.

In this paper we provide an insight into the family of systems for plagiarism detection developed by Masaryk University. Furthermore we introduce well-established procedures for preventing and dealing with plagiarism, which are used at Masaryk University and which are also recommended to other institutions using these systems. Providing anti-plagiarism services (for more than 40 institutions) offers the wide field of experiences. Last but not least the benefits of joint effort in anti-plagiarism activities will be presented.

Text below is organized as follows. In section 2, the issue of plagiarism in academic world and reaction of Masaryk University will be discussed. That also include brief introduction of systems Theses.cz, Odevzdej.cz and Repozitar.cz. Section 3 focuses on detecting similarities against Internet sources. Section 4 summarizes benefits of presented approach and finally, in section 5, we give our conclusion.

Background

The issue of plagiarism has recently been intensively discussed in academic and scientific circles. The increasing availability of online data such as learning materials, theses or scientific articles tempts students, and sometimes even teachers, to submit someone else's results as theirs. This situation calls for tools capable of revealing instances of plagiarism, which would, at the same time, function as a deterrent to the would-be copyright violators.

As a reaction for this trend, developer team of Information System of Masaryk University have created a service for plagiarism detection which was initially intended only for university purposes. Consequently, the service has become the basis of the platform on which the anti-plagiarism systems Theses.cz, Odevzdej.cz and Repozitar.cz were built. The core of this uses proprietary distributed chunk-based algorithm. From the text form of each document, there are extracted overlapping chunks of consecutive five words. The algorithm evaluates portion of the same chunks found among documents. More information about used algorithm can be found in (Kasprzak et al., 2008)

For successful fight again plagiarism it is very important to hold huge base of data. Therefore for all three systems their mutual interconnection is characteristic, thus similarity searches can be done across all files they contain. Thanks to the quality of modern search engines and the huge amount of online data it was also desirable to include other sources inclusive of the Internet, detailed description can be found in next section "Detecting similarities against Internet sources".

Thanks to the common platform developed systems can share even other useful, during years well-approved, features. They provide several options how records, consisting of meta-data and set of attachments, can be inserted into the system. That includes both manual method for individual or bulk insert by users via system-specific web forms and different forms of automatic downloading from remote systems of participated universities and other institutions. Bulk uploads of records is using xml files with specific semantic and syntax. Each system has its own set of allowed elements given by its individual purposes. The most commonly used automatic methods are based on protocols OAI-PMH or transmission of the data using the program Curl. Once the record is stored, the embedded OCR system then automatically converts the attached files into the plain text and PDF format. Typically, the PDF file serves to public presentation and text format is used for further processing including full-text search system indexing, splitting into chunks of text for finding these chunks inside other documents, and other tasks of text analysis. For participants using automatic methods the plagiarism detection task is most important and they would like to have this information available even in their own information systems. Therefore as soon as the similarities are computed they are automatically exported back to the interconnected system.

The access interface of all the systems more or less uses common full-text search engine which indexes both meta-data as well as the full text of the thesis, seminar papers, articles or other publication records. Last but not least the common communication tools should be mentioned. They include discussion forums which enable users to

share their problems, advice or experiences with other users, and message boards for mass communication.

Bellow, the description of individual systems follows, focusing on their specific characteristic.

Theses.cz

In 2006, Masaryk University provided its students and teachers a unique service that helped them to identify potentially plagiarized parts in the thesis. At that time, demand for a similar service arose at other universities as well. This demand represented an impetus for the joint project Theses.cz—National Registry of Theses and Plagiarism-Tracing System. It has started in 2008 and involves Masaryk University as well as the 16 other universities. Later on, some other universities joined the system. Nowadays there are 38 (36 Czech and 2 Slovak) universities involved in the project altogether.

The main objective of this system is to provide plagiarism detection for schools involved. Theses.cz is designed to cover different requirements. Every institution uses the system in own way according to internal procedures. The most of customers collect the theses in electronic mode via local information system. It turned out that theses' supervisors necessarily need effective tools for exploring the originality of submitted works. The system also serves as an archive of theses, with the option to publish metadata and full texts in several modes according to individual configuration. Although the recommended configuration is to publish all theses in accordance with The Higher Education Act and Open-Access principles, many of the universities involved hide the full texts. However, there are not hidden from the plagiarism detection system. Thanks to the Theses.cz, the unpublished documents enrich the corpus, too.

Odevzdej.cz

Even before the emergence of Theses.cz we held discussions with representatives of other universities on the topic of plagiarism in seminar papers. Originally, we thought that it would be possible to use Theses.cz for this kind of students' works. Due to different copyright mode of seminar papers, this showed up later as not optimal solution. In the Theses.cz there are stored metadata for each thesis and system has to solve accessibility of metadata and full texts. On the contrary, generally there is no need to collect any metadata and publish seminar papers. Therefore we came with an idea of individual system for seminar papers shortly after the introduction of Theses.cz. At that time, 9 other universities in the Czech Republic decided to participate in the project of developing system for plagiarism detection in seminar papers—Odevzdej.cz. (Brandejs, M. et al., 2009)

Odevzdej.cz provides partial e-learning solution for collecting seminar papers and essays to schools involved into the project. Teachers create so-called Vaults—the folder where students upload their papers. This eliminates the administration associated with the collection of papers, such as personal communication by e-mail and individual downloading attachments or subsequent archiving. There is a list of actual submissions in the Vault which is always available for the teacher. The system automatically closes submitting to the Vault in accordance to the deadline, which can be set in the Vault

options. Odevzdej.cz provides to teachers tools for managing, plagiarism-detecting and evaluation of uploaded papers.

Odevzdej.cz is open to anybody. Everyone can check 3 papers a day without the need to register in the system. The developer team was inspired to this approach by experiences with Theses.cz. There was a high demand for inspection of papers by public users (without an authenticated access into any of the mentioned systems). The uniqueness of such free access to the system lies in the fact that everybody has the opportunity to inspect the document for plagiarism and thus contributes to greater copyright protection. User can upload the file he wishes to check via the uploading form at the title page of the system. After processing uploaded document, the system sends results to the user-specified e-mail address. The files are automatically deleted in 5 days and nobody else can see the similarity with these temporary files.

Repozitar.cz

System Repozitar.cz has arisen in order to provide services for long-term storage and presentation of academic papers and other scientific outputs (Jakubík et al., 2011). It serves the purpose of digital library and includes technical solution from the necessary organizational, social and legal environment. According to the Open Access idea there is an effort to make the most of records open to the public but for various reasons it is not possible to fully achieve the goal. Therefore the system handles a wide range of access rights through which the authors can restrict the access to their publication outputs.

The main access interface of the system is based on a full-text search which indexes both meta-data as well as the full text of the articles and other publication records. This full-text search system is also used for searching of similar documents or generating various lists through its ability to add any additional information in the form of “virtual tokens” to the index. The advanced search options extend the application by adding the possibility of progressive refinements to the query. It allows search not only by publication own meta-data but also by departments, R&D projects or other data required for transmission to the RIV. To each one of found records, the seeker can display abstract, list of citations, attached files, similar documents and other detail information. Additionally, a final list of matching records can be entered into a user box which enables the transfer of a selection into other applications, work with it and subsequently process them en bloc.

Detecting similarities against Internet sources

The easiest way for plagiarizing is to use favorite search engine to obtain an online available document and use this document as a source of plagiarism. The task of candidate document retrieval is to identify for a certain input document relatively small set of source documents which may be plagiarized from. Those candidate documents are usually further processed using detail document comparison methods in order to discover potential plagiarism. It is up to the detail comparison method what kind of plagiarism it is able to discover. Nevertheless the standard methods use algorithms

based on document similarities which may be infeasible to compute among large collection of documents which is in case of online plagiarism the whole Web.

Since the detailed document comparison is based on computing similarities among documents which the system indexes we need to provide the system with the source documents which may have served for plagiarism. The computational complexity and storage demands of vast amount of documents prevent us from crawling and indexing the whole Web. The most straightforward method of candidate document retrieval is utilizing ideally the same search engine as the plagiarists did. Latter-day commercial search engines like Google, Yahoo or Bing are indexing the World Wide Web continuously. Despite the fact, that they cannot provide fully up to date data, we can simply rely on them, since the plagiarist could not also find a source which has not yet been indexed by that search engine. Commercial search engines do not provide access to their internal index, nor do they usually publish used retrieving methods in detail. Therefore we are using standard query interface the very same as the regular users do. The problem of candidate document retrieval is then converted into the problem of constructing befitting queries for the selected search engine. Accordingly it is wise to use more than one search engine. Our research shows that generally more search engines provide slightly different results on the basis of the same query.

We must apply several considerations on constructing queries from suspicious documents. The first consideration is to minimize the total number of executed queries, because the number of executed queries is usually limited by the search engine or not free of charge. Also every query takes not insignificant amount of time to process. Second consideration is to decide how many and which of the resulting documents to download and process which influence the performance measure. We must also maximize the precision and recall of the downloaded documents regarding actual potential plagiarism sources of the suspicious document.

There is a need to obtain only relevant documents. We define a document relevant if it is similar to the input document in the way of document similarity computed by our document similarity evaluation algorithm or if it discusses the same theme.

Two documents following the same theme usually share a set of overlapping words. Consequently querying quality keywords extracted from a given text would result in obtaining relevant documents following the same theme. Assume that an input document does not cover more than one coherent theme. That is the case of for example theses and seminar works which we mainly focus on. On the contrary for example one edition of newspapers usually contains many different articles. Based on this assumption we extract key-words from the whole document. This means that they cover the document as a whole and serve for obtaining theme-bounded documents. We developed an in-house key-word extraction algorithm currently supporting 5 European languages, which is easily extensible. The algorithm is based on term frequency analysis combined with the TF-IDF statistical measure.

For constructing queries, we are using similar approach as we used during the international competition on plagiarism detection PAN 2012 published in (Suchomel, Kasprzak & Brandejs, 2012). Our approach resulted in the achievement of the best results in overall performance of the system. More information about the course of the competition can be found in (Potthast et al., 2012). The team of IS MU has competed

in the international competition on plagiarism detection PAN three times. Firstly in 2009 with overall second place in the External Plagiarism Detection task. Secondly in 2010 with overall first place in the Plagiarism Detection task. Thirdly in 2012 with second place in the Detailed Document Comparison task and with the best performing approach in the Candidate Document Retrieval task.

As opposed to the published method, we are not using the queries based on extracted headers from suspicious document since they provide the least performance benefit. We are using two types of queries: the key-words based queries; and the intrinsic plagiarism based queries.

Since December 2010 our system checked about 300 thousand documents for plagiarism against the Internet. It is enriched by the candidate document retrieval process of more than 6 million relevant documents from aimed Internet search and about 7 million Web pages from the Wikipedia. The Wikipedia is crawled and stored on regular basis, since it is often used as a quality information source and may easily serve as a source of plagiarism.

A question we wanted to answer by means of utilizing the system is whether the Internet is in real time “inexhaustible” source of documents. In other words whether there still will be loads of new URLs found as relevant according to our search method for similar documents. Our research shows that the number of unique URLs found is significantly lowering in time of processing due to the fact, that we have downloaded lots of URLs which are repeated among searches. For example after two months of system running there were added during the next month 130 thousand new unique URLs for download and another 250 thousand were also found but rejected, because they have been already downloaded.

Plagiarism evaluation

There are several aspects of potential plagiarism that need to be considered. Percentage of similarity is only one of them. The outline of the plagiarism detection is a list of the similarity rates (percentage of similarity) linked with the source documents. Described systems provide the ability to explore each similarity by clicking the link which opens your document with highlighted suspicious similar passages. The user has to inspect the highlighted text and evaluate compliance with citation ethics and the range of citations whether they correspond to the type of paper/thesis. The problem is that some users expect from the system decision whether the paper/thesis contains plagiarism or not. After the detailed examination of all aspects (of all similarities from the list), the final evaluation could be formulated.

In theses context, there is very important the ongoing cooperation between students and their advisor. The quality of the thesis can be controlled better through incremental paragraphs than final text. The advisor can affect the content, form of the thesis and also citations. The final decision about an examined text should be the collaborative decision of Thesis Defense Committee or Disciplinary Board. Any system cannot be responsible for evaluation and following sentence whether a thesis is in accordance with requirements on the specific type of work. The advisor, opponent and Thesis Defense Committee are guarantors of the quality. They also have the option to refuse

the thesis. After the thesis has been defended, additional objections can hardly be taken into account. Therefore, it is strongly recommended to check and evaluate plagiarism by advisor (or other responsible person) before the thesis defense.

Checking of scientific articles and other works without student/advisor relation must follow different approach. Therefore on-line availability of plagiarism detection module for regular users (not only administrators, teachers or persons with “high” access rights) is characteristic for all three systems. This ensures that every document can be controlled at any time by anyone.

Even if very well similarity detection is available to users, the issue of plagiarism can never be decided by a computer system. The final verdict about plagiarism content must be done by a human; the system tries to facilitate the decision process and point out suspicious documents or suspicious document parts. (Suchomel, 2012)

Reflection of inter-university cooperation

The area of the plagiarism is very large and anti-plagiarism systems can be utilized in many different cases. The system Theses.cz can help the teacher to identify suspicious paragraphs in theses and seminar papers. Contrary, the students can benefit from public access to the Odevzdej.cz and inspect their own theses before submission. They can easily check the correctness of citations or consult the range of passages taken over from another author. Electronically available scientific papers could be abused similarly as freely accessible theses. Therefore, system Repozitar.cz combines functionality of digital library with plagiarism detection.

A significant benefit is the prevention effect of plagiarism detection systems. The risk of revelation connected with systematic usage of anti-plagiarism systems by tens of universities in the region further reduces the determination to misuse available electronic sources. To multiply the effect, it is important to raise awareness of citation ethics and copyrights. From our experience that leads to increasing quality of all types of academic works and possible improvement of the university prestige. Moreover, the obligation of publishing theses on the Internet naturally influences on the quality.

The systematic use of plagiarism detection tools requires integration to the common systems. A professional anti-plagiarism tool needs quality similarity detection algorithms which can scale up to the web. Also integration into internal guidelines and Study and Examination Regulations is needed. The students' theses, seminar papers, essays, etc. should be inspected not just randomly. Institutions should have clear rules and competencies to detect plagiarism.

Another interesting outcome on the quality of all types of works should have their publication on the Internet. People act naturally more responsibly knowing that their text will be obtainable.

Conclusion

In the age of information technologies the problem of plagiarism has become more frequent and turned into a serious issue. In this paper, we described its' complexity and the role of technology in combating it. We discussed the fact that plagiarism detection tools provide excellent service for detection of similar text between documents and they

can help users to easily identify suspicious works or paragraphs. But the final decision based on exploring the similarities and their evaluation are up to the users. They are responsible for the decision about plagiarism and they have to be prepared to submit substantiating arguments.

We provided an insight into the family of systems for revealing plagiarism which arises at Masaryk University. It consists of systems for plagiarism detection in theses (Theses.cz), in seminar papers (Odevzdej.cz), and in scientific papers (Repozitar.cz). All three systems are based on the same platform and share common database. Since the Internet can be considered as the best pool of resources for plagiarism, we have pinpointed principles and techniques which we use for detecting similarities against this vast amount of online data. Last but not least the enlightenment through seminars and workshops about plagiarism research and proper citation methodologies should be mentioned.

With such systematic approach to plagiarism, we hope that the institutional culture of the university can be transformed in a way, that the need and desire of students to plagiarise could be dramatically reduced. For the future, many tasks remain; nevertheless we can already see some great results that we have achieved so far. Together our team is preparing for the collaboration with other institutions in this area.

References

- Brandejs, M. et al., 2009. Odhalovat plagiáty se daří díky spolupráci a odpovědnosti vysokých škol (Cooperation Among Conscientious Universities Reveals Plagiarism). In *UNINFOS 2009*. Nitra: Slovak University of Agriculture in Nitra
- Jakubík, D. et al., 2011. A central repository of publication results, implemented as a part of systems for revealing plagiarism. *Seminář ke zpřístupňování šedé literatury 2011*, 4.
- Kasprzak, J. et al., 2008. Distributed System for Discovering Similar Documents. In *ICEIS 2008*. Brno: Faculty of Informatics, Masaryk University, p. 14.
- Suchomel, Š., 2012. *Systems for Online Plagiarism Detection*. Thesis. Brno: Masaryk University.
- Suchomel, Š., Kasprzak, J. & Brandejs, M., 2012. Three way search engine queries with multi-feature document comparison or plagiarism detection. In P. Forner, J. Karlgren, & C. Womser-Hacker, (eds.) *CLEF (Online Working Notes/Labs/Workshop)*. Available at: <http://dblp.uni-trier.de/db/conf/clef/clef2012w.html#SuchomelKB12>.
- Potthast, M. et al., 2012. Overview of the 4th international competition on plagiarism detection. In P. Forner, J. Karlgren, & C. Womser-Hacker, (eds.) *CLEF (Online Working Notes/Labs/Workshop)*. Available at: <http://dblp.uni-trier.de/db/conf/clef/clef2012w.html#PotthastGHKMOTBGRS12>.

Authors

Luboš Lunter, lunter@fi.muni.cz, Daniel Jakubík, jakubik@fi.muni.cz, Šimon Suchomel, suchomel@fi.muni.cz, Michal Brandejs, brandejs@fi.muni.cz, Faculty of Informatics, Masaryk University, Czech Republic

Copyright © 2013 Authors listed on page 216: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced.

The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

STUDENTS AND STAFF VOICES ON “ZU GUTTENBERG’S CASE” AND ITS INFLUENCE ON PLAGIARISM AWARENESS IN GERMAN HEIS

Anna Michalska

Abstract: Plagiarism is one of the greatest concerns in academic circles and it became even more “popular” after uncovering cases of academic dishonesty committed by high ranking politicians. It started with German defence minister, Karl-Theodor zu Guttenberg proven to copy 63% of the lines on 94% of pages in his PhD thesis without referencing the sources.

People responsible for exposing these “prominent” cases are being called “die Jäger” (eng. the “hunters”) and their reputation is spreading across Europe. Despite their noble intentions to catch and punish all the cheaters, do they positively or negatively influence German Higher Education? Does discussion about politicians’ doctoral theses help to raise awareness about the problem or does it only create atmosphere of fear and intimidation?

Impact of Policies for Plagiarism in Higher Education Across Europe (IPPHEAE) is a project that investigates equality and coherence of standards in Higher Education across the European Union. The project has discovered major disparities and gaps in policies at national and institutional level in many European countries. Being a member of the IPPHEAE project team, the author received an opportunity to focus on her own research which is an offshoot of the main project and concentrates on academic dishonesty from the student perspective.

This paper presents results of observations made during the author’s visit to four German HEIs in November 2012. Conclusions are based on opinions of 20 students and 12 members of staff collected during focus groups and interviews carried out as a part of the IPPHEAE project.

Key words: student plagiarism, German Higher Education

Setting up the scene: A few words about Karl zu Guttenberg

Karl Theodor Maria Nikolaus Johann Jacob Philipp Franz Joseph Sylvester Freiherr von und zu Guttenberg or simply Karl-Theodor zu Guttenberg was a German “rising star” politician of the Christian Social Union (CSU), who ended up being publicly humiliated and despised for his dishonest academic behaviour.

He became a Member of Parliament in 2002 where he held the post of Secretary General of the CSU party and then was appointed as Federal Minister for Economics and Technology in the first Angela Merkel cabinet on 10 February 2009. In her second cabinet, from 28 October 2009, zu Guttenberg became the Minister of Defense and one of the most popular German politicians.

In the meantime in 2006, under supervision of Peter Hübner, zu Guttenberg completed his dissertation for the University of Bayreuth and in 2007 was awarded a doctorate in law receiving the top grade of summa cum laude. His thesis was published under the title “Constitution and Constitutional Treaty: Constitutional developments in the USA and EU” (Kelsey, 17 Feb 2011).

In February 2011, zu Guttenberg was publicly accused of violating academic standards in his thesis by extensively copying information from many sources, such as newspaper

articles, public lectures or term papers, without referencing them correctly. On 18 February 2011 Guttenberg made a statement that “I will temporarily, I repeat temporarily,” “give up my doctoral title” and added that “My dissertation is not a work of plagiarism and I completely reject all such accusations” (BBC, 18 Feb 2011). However three days later, on 21 February 2011, he asked the University of Bayreuth to permanently invalidate his academic degree after admitting to make “serious mistakes” and “do not meet the ethical code of science” (BBC, 22 Feb 2011). The university retracted his degree on 23 February, based on their own investigation which provided evidence that zu Guttenberg had “violated scientific duties to a considerable extent” (BBC, 24 Feb 2011).

On 24 February 2011 Bundestagspräsident Norbert Lammert said that zu Guttenberg had used six reports created by the parliamentary research service without obtaining necessary approval. However, Lammert admitted that it was a common practice among Bundestag Members to use such documents without authorisation. The Bundestag decided not to press any charges against zu Guttenberg. Even Chancellor Angela Merkel stood behind her “rising star” saying that she “appointed Guttenberg as minister of defence (...) not an academic assistant or doctor”. Merkel also added that “What is important to me is his work as minister of defence and he carries out these duties perfectly” (BBC, 22 Feb 2011).

Both Guttenberg’s misbehaviour and Merkel’s backing raised criticism and anger in German academic circles. As a result, Chancellor received an open letter signed by over 50,000 researchers and PhD students, expressing their disapproval for Guttenberg’s continued role in her government (BBC, 6 May 2011).

Entanglement in plagiarism scandal, public humiliation and invalidation of his doctoral thesis led to Guttenberg’s resignation as Minister of Defence on 1 March and from the Bundestag on 3 March 2011. Upon further investigation, it was found out that more than half of zu Guttenberg’s 475-page thesis contained large sections intentionally plagiarised from other sources. Bayreuth University claimed that Guttenberg had “grossly violated standard research practices and in so doing deliberately deceived” (BBC, 6 May 2011).

“The hunters”: Who are they?

The analysis of zu Guttenberg’s thesis was done on the online crowd-sourcing platform GuttenPlag Wiki (de.guttenplag.wikia.com) and was followed by similar initiatives on VroniPlag Wiki (de.vroniplag.wikia.com).

When day turns to night, the online chatroom VroniPlag Wiki starts buzzing. At 6:00 p.m., the first users begin to appear on the screen. By around 10:00 p.m., the majority of them are online. They’re looking for plagiarized work in doctoral theses. They do this all through the night, after their day jobs in many cases, but they don’t get paid for it (Binder, 2012).

For some of the internet activists it is almost like a hobby to check the originality of the prominent theses (Kelsey, 26 Apr 2011). Some have had a personal experience in “copying and pasting” and joined online community after plagiarising their own doctoral dissertations. Other “hunters” started to realise the scale of the problem after becoming aware of zu Guttenberg’s case (Binder, 2011). Taking part in the plagiarism

seeking campaign is open to anyone. The results of each search are checked by a number of community members to ensure accuracy and merit. The most common text-matching tool is Google which is used to compare suspicious and unreferenced passages from dissertations with other documents accessible through the Internet in order to find similarities and matches (Naumann, 2011). As a recognition and appreciation of the good work (documenting and discussing zu Guttenberg’s case as well rising public debate about plagiarism) the GuttenPlag Wiki and its all contributors received in 2011 the Grimme Online Award in the “Special” category (http://de.guttenplag.wikia.com/wiki/Grimme_Online_Award).

Despite their noble actions, “the plagiarists’ hunters” have to face many criticisms. In August 2011, the authors of the online platforms were accused of running a political campaign against CSU party after revealing that the founder of the VroniPlag, and also an active affiliate of the GuttenPlag Wiki, is a member of the opposition SPD (social democratic) party (Badische Zeitung, 2011). Another accusation concerned their identities. Some observers believe that despite uncovering academic dishonesty and triggering positive actions, community of plagiarism hunters hides behind the online anonymity. Only a few decide to make their real names public. However they have their motives. Investigating prominent figures and going after colleagues is seen as betrayal and therefore the hunters with a career in academia prefer to remain anonymous (Binder, 2012). Members of the online community feel threatened. They understand that a person who uncovers such scandals is often seen as a whistle-blower and for that reason they decide not to put their careers at risk. They believe that the pseudonym can offer them a reasonable security (Naumann, 2011).

Anonymity of the online community has raised many questions concerning ethics of their actions. The biggest would be damage to reputation in case the allegations turn out to be groundless. Some observers agree that the online crusade is a positive initiative and it is important that society starts to pay attention to plagiarism problem, though it may appear as an uneven fight when the alleged plagiarist does not know who stands behind the accusation (Kelsey, 26 Apr 2011).

One of “the hunters”, Debora Weber-Wulff, a professor of media and computing at the HTW science university in Berlin, explains anonymity of activists by highlighting the fact that it is not about the attacker, but about the issue. She disagrees that their actions can irreversibly damage reputation of the alleged person: “They’re not accusing anyone of plagiarism, they’re saying, ‘Can you explain why no quotes were used?’” (Kelsey, 26 Apr 2011). What is more, Weber-Wulff believes that the problem lays with lack of universities’ reaction to the academic dishonesty issue: “As long as universities fail to take measures against plagiarism, many plagiarist hunters will remain anonymous” (Binder, 2012). She has a strong opinion about condition of Germany Higher Education claiming that many cases of academic misconduct have been “sweeping under the table” and that the situation got out of control (Kelsey, 26 Apr 2011). Debora Weber-Wulff frequently and passionately highlights her belief that there is no real support for plagiarism prevention and detection at German universities and the lack of training for tutors and teachers only worsens the situation. Moreover there are no procedures for dealing with plagiarism cases at lower levels of education.

Zu Guttenberg's case was just a starting point. According to Debora Webber-Wulff and Graf Isolan (2012):

Currently 26 cases are documented on the site. Of these, eight doctorates have been rescinded (with several lawsuits pending); three have been declared to be within the bounds of acceptability by the awarding universities, although those institutions have provided no explanations for the substantial numbers of text parallels. (...) Other cases not on VroniPlag Wiki have involved the Romanian minister of education [Ioan Mang], the Romanian prime minister [Victor Ponta], the Hungarian president [Pal Schmitt], an official in Thailand, and a parliamentarian in South Korea. Documentation is also underway in Russia concerning the dissertation of their new education minister [Vladimir Medinsky].

The authors also added "The extensive documentation has demonstrated that plagiarism is not just an occasional incident, but something that the German university system must now get serious about" (Webber-Wulff and Isolan, 2012).

Another prominent figure whose case hit the headlines was German Education Minister Annette Schavan accused of plagiarising parts of her doctoral thesis in 1980. Investigators discovered over 60 cases of paraphrased passages used without references. Mrs Schavan was found guilty of "intentional deception through plagiarism" and her PhD title was revoked (Diehl and Trenkamp, 2013). On 9 February 2013 she announced her resignation as minister.

Research: Assumptions of negative impact

The author of this paper is a research assistant to the Europe-wide plagiarism project, Impact of Policies for Plagiarism in Higher Education Across Europe (IPPHEAE) led by Irene Glendinning from Coventry University. The project is a cooperative work of five partners:

- Coventry University, United Kingdom
- Lodz University of Technology, Poland
- Mendel University in Brno, Czech Republic
- University of Nicosia, Cyprus, and
- Aleksandras Stulginskis University, Lithuania.

IPPHEAE aimed to conduct a survey in all 27 European Union countries at three different levels: management responsible for setting out plagiarism policies; teaching staff that deal with plagiarism cases according to established regulations; and students who try not to be tempted by available "shortcuts". IPPHEAE team also carried out a set of interviews with national representatives who were able to present an overview of the plagiarism situation in their country.

IPPHEAE project focused on investigating participants' understanding of academic dishonesty, as well as perception of plagiarism policies within institutions and countrywide. As a member of the team the author received an opportunity to access the project data and its participants to conduct her own research. The aim was to carry out a comparative study of different European nationalities and find out whether students

from diverse countries and backgrounds present common or dissimilar views towards plagiarism problem. As a part of her PhD, she conducted a set of focus groups that brought new dimensions and qualitative data to the IPPHEAE project.

Trying to approach many German HE Institutions and national authorities, IPPHEAE team was faced with many negative responses from potential participants who refused to be interviewed and talk about plagiarism issues in their country. The team was aware of “zu Guttenberg’s case” and suspected it may have been linked with the reluctance of some people to be interviewed. The project team believed that an atmosphere of fear and intimidation may have been created as a result of “the hunters”. The researchers began highlighting the positive purpose of their research with the idea of sharing the good practice and improving academic standards. IPPHEAE finally managed to receive some positive feedback and conducted 47 student, 7 teaching staff and 2 senior management surveys in German HEIs. The team also interviewed 14 national and university figures (senior academics, university leaders, managers, research integrity leaders) from across the sector. The author was able to carry out 5 focus groups and interviews with 21 master students during her visit to Germany in November 2012. The findings of this research are presented in the next chapter.

All data collected in the survey and during focus groups was held anonymously and securely in order to ensure that none of the responses would lead to the identification of a participant or institution. Codes and details of participants whose views and opinions were used in this paper are shown in a table below (contribution of some participants was omitted as their comments were irrelevant to this paper).

Findings: Positives and negatives

The plagiarism situation in Germany can be best described in the words of the interviewed participants. They have mentioned many negative, but also positive characteristics of German Higher Education. What is more, they recommended some actions to improve the system and implement constructive changes.

Starting from negative comments stated by national and university representatives, many respondents mentioned general lack of awareness and discussion about plagiarism problem:

- (Nat40) “There is no great awareness.”
- (Nat16) “There’s little discussion.”
- (Nat16) “‘None issue’ in Germany, each professor is responsible for his own problems”
- (Nat20) “*More training for students and staff?* Absolutely—very few. (...) We do not do enough for quality of teaching and learning, too many student per instructor, no incentives for good teaching.”

Interviewees admitted increase in number of plagiarism cases:

- (Nat20) “In the last few months 2 or 3 surveys have been completed in Germany, but not published. Consistently showing up to 40% of students regularly copy and paste content from sources without citing and referencing. This is a big issue in

Table 1

Participants' codes and details

Participant Code	Survey Level	Participant Description	Survey Type	Institution	Number of Participants
Nat16	National	Plagiarism expert, academic	Interview	University E	1
Nat20	National	Plagiarism expert	Interview	N/A	1
Nat34	National	Academic and researcher	Interview	University F	1
Nat40	National	Academic	Interview	University G	1
Nat41	National	University Rector	Interview	University B	1
Nat42	National	Founding Rector	Interview	University D	1
Nat43, 44	National	Chairmen of Board of Examiners	Interview	University D	2
Nat45, 46	National	Academics	Interview	University D	2
Nat47, 48	National	#47 Exam Officer #48 Associate Dean of Academic Services	Interview	University C	2
StuA	Student	Master students	Focus Group	University A	8
StuB	Student	Master students	Focus Group	University B	5
StuC.1	Student	Master students	Focus Group	University C	6
StuC.2	Student	Former master student	Interview	University C	1

secondary education, known to be a general phenomenon that homework is always copied.”

- (Nat41) “[Plagiarism is] doubtlessly increasing, even professors would agree to this.”

Unfortunately not many statistics are available:

- (Nat43, 44) “There are statistics at the level of institutes—not faculty or university level.”
- (Nat45, 46) “We keep statistics for departments, but there will be huge difference between real cases and documented ones. (...) The system is very draconian, so some colleagues prefer not to record cases. It’s a structural problem—the more severe cases go to a court, where is a different set of penalties, but if cases are not severe many professors don’t want to bother.”
- (Nat16) “No statistics whatsoever therefore would not know in general, however we find more cases—maybe because we are looking better.”

National and university representatives stated many interesting aspects of German plagiarism culture:

- (Nat16) “It is about attitude, a culture of plagiarism has grown up”.

- (Nat34) “Some [lecturers] don’t know, some deliberately ignore, don’t want people to question their practices, and some chase everywhere, strict leading to exclusion of students”.
- (Nat45, 46) “There is no consistent approach towards plagiarism. Some people don’t want to be ‘controllers’—they are here for different reasons”.
- (Nat16) “There is no transparency in Higher Education in Germany”

Participants also mentioned lack of national as well as institution-wide policies:

- (Nat16) “I have a slide in my presentations of an ostrich with its head in the sand that describes the attitudes in Germany. No national or state policy, although some states may set some guidelines, many don’t”.
- (Nat20) “In Nord Westphalia rectors have recommended policy for using software to detect plagiarism. But many academics think we cannot put students and scientists under general suspicion—should only act where there is strong suspicion of cheating”.
- (Nat20) “Very few institutions have top-down strategy/policy for preventing and detecting plagiarism. Highly decentralised”.
- (Nat16) “No evidence of procedures being effective.”
- (Nat20) “There are no institution-wide policies therefore can’t be effective”

Fortunately there were many positive aspects stated by the interviewees:

- (Nat45, 46) “We are getting more precise and detailed about our policies. We have list of penalties for each case.”
- (Nat16) “Some institutions are purchasing software”
- (Nat45, 46) “On the first semester of undergrads—they [students] do special course about academic writing—different aspects of politics and social science, but it’s about testing their awareness.”
- (Nat42) “[Policies] somehow effective, explaining, getting aware of plagiarism. More attentive. But we could do more.”

Participants had many suggestions and recommendations on how to improve the system:

- (Nat20) “Trying to detect is the wrong way—Universities should start to invest in infrastructure.”
- (Nat43, 44) “The key point is to give students an assignment which they cannot plagiarise.”
- (Nat20) “Needs national debate and discussion. It should be banned. Every university should develop policy, clear communication to students and professors about it. Invest in infrastructure and provide lessons to help students to be better writers of homework, thesis. Universities need to reduce the number of students per teacher.”
- (Nat41) “Talking about it, explaining, showing examples, trying to know, it is a question of honesty, students should be proud to be in academia.”

- (Nat41) “If you work with students during writing of their thesis there will be less plagiarism in the final version.”
- (Nat34) “Teacher dialogue, bring into open and talk about it more.”
- (Nat16) “Training, transparency, testing—this is what Germany needs!”

What is more, during interviews with representatives of German HEIs, the team has found out that all of the interviewees were aware of “zu Guttenberg’s case” and recognised it as a step towards improving anti-plagiarism system in Germany:

- (Nat43, 44) “There is a psychological effect after zu Guttenberg’s case.”
- (Nat42) “Starting point was zu Guttenberg’s case—it raised public interest and awareness at institutional level.”
- (Nat41) “There has been a changed view. A year ago no one was talking about it, but because of zu Gutenberg people are now aware, attention has increased.”
- (Nat45, 46) “Guttenberg and other cases had a major influence on PhD level studies. Reactions were varied and different questions were asked: How to raise awareness? Should we do electronic checks? Etc.”

Research participants mentioned exposing prominent cases as a trigger for positive actions across Germany and listed some new initiatives:

- (Nat41) It has been discussed at the state level. There is a recent policy paper for all Universities of Applied Science, decided 6 months ago. The new policy about safeguarding good academic practice contains several new rules and regulations, such as:
 - Explain to students what plagiarism is
 - All final theses should be kept in electronic copies
 - Faculty/Department defines the person to whom thesis given
 - Confidentiality clauses
 - All final theses must be checked for plagiarism
 - If evidence of plagiarism found, thesis has to be redone.
- (Nat41) “Confederation of Vive-Pro Rectors for teaching has been discussing this—whether to be forced by law to do some evaluation, peer-review.”
- (Nat16) “Suggestions were presented to the Bundestag, recommended 3 pronged programme:
 - 1: Educate people—set up a central body to educate teachers about how to avoid plagiarism
 - 2: Transparency, open access, digital submission of dissertations—will expose some shocking examples
 - 3. Mode of controlling—federal body chooses a sample from last 5 years to identify where the problems are, introduce a QA process.”

There were some other aspects raised when talking about “zu Guttenberg’s case”:

- (Nat47, 48) “Guttenberg plagiarised his supervisor—why did the supervisor not find it? Perhaps experts do not read the thesis?”

- (Nat34) “Guttenberg, nobleman (...)—no one would question honesty.”
- (Nat20) “The whole culture starting with secondary education is far too soft. I want to point out that the first public Minister, zu Guttenberg. The Chancellor Mrs Merkel, who has PhD herself, was asking ‘why are you so critical, he only copied, it is not important! I hired him as a minister not a scientist.’”
- (Nat19) “Public cases have been exposed, but not from institutions about cases not so prominent—ministers etc., but behind that I think there is a lot going on.”
- (Nat16) “We expose people and in the press we’re nasty people, not those who have plagiarised.”

Some participants expressed their support towards “the hunters”:

- (Nat40) “*Examples of good practice?* Debbie! [Debora Webber-Wulff]”
- (Nat40) “Why does VroniPlag only exist in Germany?”

Despite giving many recommendations on how to improve German anti-plagiarism system, many teaching staff participants did not recognise existence of the problem among themselves, inside academic circles. When asked “*Do you think there should be more training about preventing plagiarism and academic dishonesty for staff?*” some of them answered:

- (Nat42) “I don’t think they [staff] need training—maybe in a form of open discussion with students. Academics know about plagiarism.”
- (Nat45, 46) “It would be ridiculous to teach staff about plagiarism, professors are already experts and authorities when they are appointed to their posts.”
- (Nat16) “German people link plagiarism with ‘copyrights’, so they think it’s ok to use your own words, ‘self-plagiarise’, as you have the copyrights.”
- (Nat47, 48) “Self-plagiarising—not just one view—some people say it does not exist, i.e. it is still my own idea, no need to reference.”
- (Nat45, 46) “Self-plagiarism is OK at PhD level.”

Many sarcastic comments from press only exacerbate the atmosphere calling zu Guttenberg “Mr zu Googleberg” or “Baron Cut-And-Paste”. On the other hand public interest in plagiarism matter is only visible when journalists expose a case involving a prominent figure. Publicising plagiarism affairs helped increase student awareness about academic integrity. Students participating in the research said:

- (StuB) “We’ve learned about plagiarism from the news!”
- (StuA) “We don’t know many [plagiarism] cases personally, but many nationally.”

Most of the interviewed students admitted that revealing “zu Gutenberg’s case” influenced the quality system at many universities and made the rules more rigorous:

- (StuA) “Since Guttenberg it gets really strict”
- (StuB) “F***** Guttenberg!”

Despite some negative comments from the students who obviously do not like the fact that the rules are now stricter and they have to behave more properly, many students are happy about new regulations and see it as a chance of “healing” the system.

At one of the institutions where anti-plagiarism policies were already implemented and used, students had mixed opinions about impact of “zu Guttenberg’s case” on teaching and learning style at their university:

- (StuC.1) “We don’t see any difference after zu Guttenberg’s case”
- (StuC.2) “Zu Guttenberg had influence—it damaged the reputation of PhD students! Our university is quite rigorous, but it could have influence.”

Conclusions: What’s next?

IPPHEAE project team managed to reach only a few HEIs in Germany. These cannot represent the whole German education system. It is possible that only “good universities” which are more concerned about plagiarism issues than others decided to share their experience and spread positive actions across academia. It is also probable that some universities which remained silent are ashamed of their poor approach towards academic dishonesty or are simply unaware of gaps in their policies and procedures. Taking into consideration all the problems the team had to challenge, as well as some negative comments from survey participants, it is reasonable to conclude that there is a fear that all “plagiarism projects” have only one thing in mind—to catch the cheaters. Maybe we all have something to answer for.

Guttenberg’s guilt is undeniable. But what with other plagiarists like Annette Schavan? Should we punish for crimes conducted in the distant past? Where do we draw the line? According to one of the participants (Nat42) “there is nothing wrong in finding these [prominent] cases. If you did something wrong, you need to be punished”.

Despite the fact that anonymity of “the hunters” raised debate about ethics and morality of their actions, the positive influence of their campaign is unquestionable. It affected large and small universities in many subject areas and levels of study. It initiated discussion about quality of academic research and will almost certainly lead to improving standards in German Higher Education. Some institutions have already implemented new policies and anti-plagiarism regulations. Zu Guttenberg’s case “also influenced wider community while its “fame” spread to other European countries. The next step should be to introduce positive initiatives for master and bachelor students and possibly start academic integrity campaigns at lower levels of school education.”

“Training, transparency, testing—this is what Germany needs!” says Debora Webber-Wulff.

List of References

- Badische Zeitung (2011) Plagiatsjäger enttarnt—Vorwurf der Parteilichkeit. *Badische Zeitung* [online] 5 August. Available from <http://www.badische-zeitung.de/deutschland-1/plagiatsjaeger-enttarnt-vorwurf-der-parteilichkeit--48254663.html> [10 March 2013]
- BBC (2011) German ‘plagiarism’ minister Guttenberg drops doctorate. *BBC News Europe* [online] 18 February. Available from <http://www.bbc.co.uk/news/world-europe-12504347> [8 February 2013]

- BBC (2011) German minister gives up doctorate after plagiarism row. *BBC News Europe* [online] 22 February. Available from <http://www.bbc.co.uk/news/world-europe-12532877> [8 February 2013]
- BBC (2011) German minister loses doctorate after plagiarism row. *BBC News Europe* [online] 24 February. Available from <http://www.bbc.co.uk/news/world-europe-12566502> [8 February 2013]
- BBC (2011) Germany’s Guttenberg ‘deliberately’ plagiarised. *BBC News Europe* [online] 6 May. Available from <http://www.bbc.co.uk/news/world-europe-13310042> [10 February 2013]
- Binder, A. (2012) Plagiarist hunters defend academic standards. *DW.DE* [online] 20 October. Available from <http://www.dw.de/plagiarist-hunters-defend-academic-standards/a-16318549> [20 February 2013]
- Diehl, J. and Trenkamp, O. (2013) Plagiarism Accusations: Merkel’s Education Minister Has Ph.D. Title Revoked. *Spiegel Online* [online] 6 February. Available from <http://www.spiegel.de/international/germany/education-minister-schavan-has-ph-d-revoked-in-plagiarism-scandal-a-881707.html> [19 March 2013]
- Kelsey, E. (2011) German minister cancels speech amid plagiarism scandal. *Reuters* [online] 17 February. Available from <http://www.reuters.com/article/2011/02/17/us-germany-minister-plagiarism-idUSTRE71G6C720110217> [9 February 2013]
- Kelsey, E. (2011) “Plagiarism hunters” plague German politicians. *Reuters* [online] 26 April. Available from <http://www.reuters.com/article/2011/04/26/us-germany-plagiarism-internet-idUSTRE73P3Z320110426> [25 February 2013]
- Naumann, N. (2011) Online plagiarism hunters track doctoral frauds. *DW.DE* [online] 17 May. Available from <http://www.dw.de/online-plagiarism-hunters-track-doctoral-frauds/a-15078344> [21 February 2013]
- Weber-Wulff, D. and Isolan, G. (2012) Uncovering scientific plagiarism. *Wikipedia Signpost* [online] 2 July. Available from http://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2012-07-02/Analysis [10 March 2013]

Author

Anna Michalska, aa8789@coventry.ac.uk, Coventry University, United Kingdom

Copyright © 2013 Author Anna Michalska grants to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the author.

PLAGIARISM IN LITHUANIAN ACADEMIA: FORMAL DEFINITION AND INFORMAL ATTITUDE

Aurelija Novelskaitė, Raminta Pučetaitė

Abstract: The paper presents some research results derived from a national project “Scientific research ethics in Lithuania: analysis of the situation” (MIP-37/2010, 2010–2011). It aims to identify gaps between the formal definition of plagiarism and actual understanding of the phenomenon in Lithuanian academic community. Therefore, content analysis of the definitions of plagiarism in approved Ethic Codes ($N = 21$) of Lithuanian science institutions and the academic community’s (ranging from professors and principal researchers to lecturers and technical assistants ($n = 424$)) attitudes concerning plagiarism expressed in a questionnaire survey was carried out. Ethic Codes were analysed using a semantic form of “plagiarism” as a coding unit. Results of this analysis demonstrate that discursive presentation of the phenomenon is more frequently related to students’ academic activities rather than academics’ research. Moreover, the Codes do not give clear-set directions how to avoid plagiarism and, thus, imply very little about plagiarism management at both the level of students and academics, although the consequences of identified plagiarism are much more clearly defined to the former than the latter. The results of content analysis of the answers to an open-ended question about the main research ethics problems observed in Lithuanian academia suggest that plagiarism can be regarded as an issue with a broad understanding, including copyright, forced authorship inclusion etc. However, the data are not sufficient to make a conclusion to what extent it is problematic and whether and to what extent consequences of the phenomenon are perceived. This calls for further research into the issue and a need for open discussion on it in the academic society.

Introduction

A query of the term “plagiarism” in a database of scholarly papers such as EBSCO yields over 3000 items. The papers cover a vast field of plagiarism issues, from the ones of “what was borrowed and never paid back” or simply copied in deep belief of following a proper tradition in *belles-lettres* (see e.g. Goodale, 1938; Masterson, 1940; Parker, 1945; Chester, 1949; Furtado, 1950; Jameson, 1993 etc.) to the ones focusing on contemporary challenges faced by journal editors, students’ plagiarism and its preventive means in the context of ICT development (Elmore, 2010; den Ouden and van Wijk, 2011; Insley, 2011).

Plagiarism, which is defined as “the presentation of the work of another person as your own ideas or intellectual property” (Elmore, 2010: 20) in this paper, constitutes one of the key forms of scientific misconduct. With a variety of twists such as paraphrasing, ghost-writing, forgetting the quotation marks, obscuring the sources etc. (Martin, 1994; Liles and Rozalski, 2004), plagiarism is reported to be one of the most widely spread violations of academic ethics among students (see Honig and Bedi, 2012) and seemingly one of the best known to the public due to, e.g. extensive media coverage on recently revoked doctorates from Hungary’s president Pal Schmitt in 2012 or the German Defence Minister Karl-Theodor zu Guttenberg in 2011 (who later resigned from their positions).

Plagiarism is not only an academic issue. It concerns public interest at large. As any other major research misconduct, it discredits the acknowledgements given by higher education institutions to their graduates, diminishes public trust in professional qualifications and social institutions in general. Moreover, cases of scholars' plagiarism undermine reputation of the researcher, weaken universities' chances to attract external funding and waste resources for research (cf. Hudson, 2008). Plagiarism hurts the plagiarised author, deceives a reader and gives undeserved merits to the plagiarist (Bouville 2008: 315–317), which can incite society's feeling of social injustice and, in radical cases, cynicism and alienation among its members. Hence, many academic institutions have made attempts to define research misconduct in honour codes, set policies and procedures for dealing with research misconduct, raise awareness of plagiarism by special courses and tutorials, introduce plagiarism detection software, establish academic ethics committees to both reduce cases of plagiarism and foster research integrity alongside with institutional excellence (Maurer et al., 2006). Yet, as Cole and McCabe (1996) note, these measures sometimes may account for confusion, concluding that "an honor code is an important, but not essential, tool in promoting academic integrity". In this respect, social cognitive theory or social learning theory (Bandura, 2009/1986) may be more explanative: observation of others' behaviour, lack or presence of articulated standards and weak or strong self-regulation and self-efficacy can respectively determine whether research integrity is upheld or transgressed. McCabe and Trevino's (1993) study has identified that "the perception of peers' behaviour was the most influential contextual variable, suggesting that social learning theory may be particularly useful for understanding academic dishonesty behaviour among college students" (p. 533).

This conclusion parallels researches by Kaptein and Schwartz (2008) on effectiveness of business codes in business context. Their research has yielded conflicting results, which raises a question about the role of the context in studying plagiarism and its causes. Causes for plagiarism have been in the researchers' focus for several decades. Yet, as Honig and Bedi (2012), Elliott et al. (2013) and others point out, the major body of this research is centred on cases of students' plagiarism while attention to academics' plagiarism is rather scarce and knowledge in this field is missing.

Another issue that we aim to address in this paper is the importance of a contextual perspective, particularly, a cultural one, in researching plagiarism. There is a body of research on the impact of the cultural background of students and their engagement in plagiarism behaviour (for some examples see Martinet al., 2009; Sisti, 2007). For example, Martin (2012) challenged a prevalent assumption that, e.g. Asian students tend to plagiarise more often than Western students because of the differences between cultures of collectivism and individualism. His study provides evidence that popular expectations that Asian students are more prone to replicating rather than producing innovative ideas based on educational material because of the importance of common norms, respect and (consequently) fear to contradict an elderly person (or one higher in social or institutional hierarchy) were ungrounded and stereotypes-based. According to Pupovac et al. (2008), "[p]revalence of plagiarism largely depends on the cultural characteristics of the academic setting and the degree to which plagiarism is implicitly allowed or explicitly accepted in the academic community and wider society" (p.

14). They proceed that post-communist countries which typically have a high rate of corruption and are “characterized by a high level of tolerance toward cheating” (ibid.) as well as lack of individual responsibility can have a higher rate of violations of research ethics. Although this idea is supported by other similar studies (Ivanauskas, 2006; Pučėtaite and Lämsä, 2008; Ryan, 2006; Vasiljevienė and Freitakienė, 2002; Žiliukaitė et al., 2006) it is little empirically evidenced. Therefore, in this paper we take Lithuania as a case of a post-soviet context and attempt to bring some knowledge on the factual situation concerning plagiarism. In Lithuania this is the first study of its kind.

To sum up, this paper attempts to make a contribution to the academic discourse on plagiarism in several respects. First, it fills the knowledge gap about the understanding of plagiarism in a Lithuanian academic society as a post-soviet one. Second, it sheds some light on the content of academic ethics codes in Lithuanian universities and research institutes as measures of managing plagiarism. Third, it aims to compare the rhetoric of research ethics codes and academia’s awareness of the issue, which can be regarded as a factual reflection of the rhetoric.

Methodological notes

The paper presents the results from a national project “Scientific research ethics in Lithuania: analysis of the situation” (MIP-37/2010, 2010–2011) which have been presented in several other publications (Novelskaitė and Pučėtaite, 2012; Novelskaitė and Pučėtaite, 2011; Novelskaitė et al., 2011). The project was based on several empirical studies: descriptive analysis of the EU documents regulating research ethics issues and the related Lithuanian legal regulations, a questionnaire survey in Lithuanian academic community, descriptive and content analysis of internal organizational documents regulating research ethics at Lithuanian scientific organizations and a series of semi-structured interviews with members of the ethics committees acting at the scientific organizations.

In this paper, for the purposes of revealing specifics of a formal definition and informal perceptions of plagiarism in Lithuanian academic community, two of its aspects were taken up for a more detailed scrutiny. First, we focused on the definition of plagiarism in organizational documents of Lithuanian academic organizations. Definitions were treated as the formal aspect of the issue. To identify to what extent formal measures are used to deepen academia’s awareness of plagiarism (and its forms) and encourage them to abstain from plagiarism conduct, Ethic Codes of the universities and the research institutes ($N = 21$; see Table 1 for details) were analysed. Analysis was carried out at the end of April 2011.

The content analysis was accomplished in several stages. Initially different parts of speech with the root *plagia** (**-rize*, **-rist*, **-rism*) (the corresponding words in the Lithuanian language are *plagi*-juoti*, *plagi*-atorius*, *plagi*-atas*) were used as units of analysis. As a result, quantitative characteristics of the issue were identified. Next, a sentence involving the unit (i.e. *plagia**-) found on the first stage was considered as the second unit of analysis and, following that, the third unit of analysis was a paragraph. The last two steps of analysis were aimed at identification of how the term “plagiarism” is defined and in what context it is presented for the target group(s) of the Code. Thus,

Table 1

Ethic Codes of Lithuanian science institutions, 2011

Type of the university*	Ethic codes**					Size/text length (No of characters with spaces)		
	N	Approved	Under construction	Non-existent	N.i.	Min	Max	Average
1. State universities	14	14***	1	—	—	3 269	17 239	9 834
2. Non-state universities	9	3****	—	5	—	6 711	16 400	12 934
3. State scientific research institutes	12	4	3	4	1	6 138	11 917	9 770
Total	35	21	4	9	1	3 269	17 239	10 265

* According to Lithuanian *Law on Higher Education and Research* (2009-04-30 No. XI-242), “there are two groups of higher education and research institutions: higher education institutions and research institutes” (Ch. II.5). More specifically, these are universities (“The university shall carry out university studies, conduct research, experimental (social, cultural) development and/or develop high-level professional art” (Ch. II.8.1)), colleges (“The college shall carry out college studies, develop applied research and/or professional art” (Ch. II.9.1)) and research institutes (“A state research institute shall conduct long-term research and experimental (social, cultural) development, important for the State, the public or economic entities, in the area defined by the founder (members of a legal person)” (Ch. II.10.2)). At the time of the analysis, *The information system of Lithuanian Education and Science Ministry* (http://www.aikos.smm.lt/aikos/svietimo_ir_mokslo_institucijos.htm) provided a list of 14 state universities and 9 non-state (i.e. public establishments) ones, 24 colleges, 12 state scientific research institutes and 2 state science establishments and 4 private scientific research establishments. For the aims of the study the list of organizations was appended by other 33 organizations, which were found in *The Catalogue of Lithuanian Enterprises: Lithuanian Yellow Pages* (<http://www.imones.lt> (accessed on August 30, 2011, keywords used in the query: “Fields of activity: scientific research, institutes”, 75 entries found) as organizations accomplishing scientific research activities and by 37 professional associations of scientists and researchers from different fields of science. (The list was compiled from different sources on internet: 30 organizations were found following links from Lithuanian Scientists’ Association (http://www.lms.lt/?q=lt/mokslo_draugijos), 6 organizations uniting professional researchers from different fields of social sciences (sociology, psychology, political science, economics, philosophy and theology) were found on internet with definite search words. From these, 35 associations (http://www.emedicina.lt/index.php?s_id=25&lang=lt) uniting medical professionals (including scientists) were skipped from the list after considering legal developments of research ethics in the field.) Thus, the project sample encompassed 134 organizations in total. For the purpose of the herein presented analysis only academic establishments providing study programmes at first-cycle, second-cycle and tertiary levels are included.

It is worth mentioning that, according to the Order of Minister of Education and Science (2005-12-05, Nr. ISAK-2485), higher education schools had to develop and pass lecturers’ ethics codes by March 1, 2006 (Ch. 2).

** Data for June 2011.

*** Lithuanian University of Health Sciences still had 2 Ethic codes—one of the former Kaunas University of Medicine and the other of the former Lithuanian Academy of Veterinary.

**** LCC International University provided a document under the title “Elements of Informed Consent, Academic Integrity and Discipline Policy”, which was not adequate to regard it as an organizational Ethic code; this document was not included into the analysis.

not only quantitative, but also qualitative analysis of a formal definition of plagiarism was accomplished.

Secondly, individual wording and descriptions of plagiarism (among other problems related to research ethics) provided by members of Lithuanian academic community as answers to an open-ended question “In your opinion, what are the most significant research ethics problems in Lithuanian scientific community and in the entire Lithuanian science system?” are taken as an informal attitude towards plagiarism. More specifically (and similarly to the content analysis of Ethic Codes), content analysis of the records expressing an opinion ($N = 913$; $n_{\text{answered}} = 424$ (46%)) was accomplished in two steps: first, the words with grammatical root *plagia*-* as the primary analysis units were identified and counted; second, taking messages that include the notion of plagiarism as the secondary units of analysis, context of the reference to plagiarism was analysed. Again, both quantitative and qualitative analysis of informal attitudes towards plagiarism was accomplished.

Results 1: Formal definition

In quantitative terms, the number of words with the root *plag**- is rather low in the analysed Ethic codes: it varies from 0 (found in the Codes of 3 state universities and 1 non-state university) to 5 (2 state universities and 1 state research institute); most frequently the word was mentioned once (in the codes of 4 state universities and 3 research institutes) or twice (in 4 state universities and in 1 non-state university) in the text of the code. It is not surprising that the number of references to plagiarism relates to the entire size of the code: the longer the Code, the more numerous mentioning of plagiarism.

Several definitions of plagiarism can be found in the codes. The most frequent one (8 cases) appears under the title “Ethics of Scientific Research (And Art) Activities” with the grounding statement running: “scientific activity must be grounded on ideals of fair research and striving for truth”. It defines plagiarism as “presentation of others’ text, ideas or discoveries like one’s own” and calls it one of the violations of the fundamental ethical principles (in addition to fabrication of empirical data, forced authorship etc.). Such a definition is presented in codes of 6 state universities and in code of 1 non-state university and 1 state research institute. Only one code of a state research institute presents a different definition under the same structure on the entire text: plagiarism is “presentation of text, data, research methodology, idea or discovery published by others as one’s own and falsification of scientific data or their unfair interpretation”.

A slightly different definition of plagiarism is found in the list of violations of the “respect for intellectual property” principle which stands for academic fairness. This definition suggests that plagiarism is “falsification or unfair interpretation of scientific data, forced inclusion of co-authorship to junior colleagues or subordinates”. Such a definition, connecting broader copyright issues with plagiarism, is presented in 1 code of a state research institute and 2 codes of state universities. Finally, still one code of a state research institute defines plagiarism as “publication of scientific data without providing references to the sources is treated as plagiarism” under the section of copyright ethics issues.

As the organization's code of ethics applies to all members of the organization (but for several cases where just particular groups are targeted by the code), all the above presented definitions should be known and followed by everyone in the organization (i.e. scientific, pedagogical and administrative staff as well as students). However, 5 codes give particular attention to students providing an additional definition to the given above under the title "Norms of students'/studies' ethics". Here plagiarism is defined as "presentation of others' ideas as one's own" and illustrated by several "typical cases", which are: (a) Cases when others' text is presented without marks of citation—quotation-marks or any other kind of segregation from the entire text (e.g. in a separate paragraph, italics); (b) Cases when others' idea, illustrative material or data are paraphrased or cited without providing an exact source; (c) Cases when knowingly a wrong page of the source or date of a web-site accession is presented. To be precise, several additional details should be mentioned. First, the last "typical case" (i.e. (c)) is provided only in 1 code (out of 5 mentioned here). It is obvious that the case is targeted at avoiding mistakes rather than plagiarism. Second, the definition and the cases are presented as one of "the crudest violations of the principle of academic fairness" in 3 codes, meanwhile self-obligation "not to plagiarise" (i.e. not to engage in "typical" plagiarism activities) is named in the 2 remaining ones. Third, the analysed 5 codes belong to 1 state research institute, to 1 non-state university and to 3 state universities. Finally, students are self-obliged "not to provide help to other persons who are engaging in an unfair academic action: plagiarism, cheating or falsification" (5 codes). Besides, lecturers (and researchers, PhD students' supervisors) "must show a principled reaction to the cases of students' unfairness such as plagiarism, cribbing, falsification of data,..." and others (10 codes), meanwhile the entire community is obliged "not to permit appearance of plagiarism, data falsification, ... etc. ... in their personal and their apprentices' works" (2 codes) as well as "not to fabricate nor manipulate empirical research data, scientific studies, plagiarize nor cheat in other ways in scientific activities" (2 codes).

The findings demonstrate several things. First, although the analysed codes slightly differ in the formal definition of plagiarism and related issues (e.g. supervision), there is no obvious difference between state and non-state universities, between universities and research institutes. These findings are not original: Forster with colleagues (2009) found "substantial levels of common sentences used by the firms, including a few cases where the codes of ethics are essentially identical" (p. 129, 138) in the compared codes of ethics developed by 597 companies till 2008.

Second, more or less precise definitions of plagiarism are provided exclusively in chapters related to research and students' activities. This finding denotes that several niches of plagiarism are not covered by the Codes (for example, plagiarism in teaching materials). This breach is especially important in the context of plagiarism prevention and academic socialization in general as teaching materials are the ones which every student faces first getting into academia.

Third, the analysis shows that plagiarism prevention is tackled by the Codes either anonymously (i.e. by an entire community) or giving particular attention to a specific group, i.e. students. This finding echoes the above mentioned remarks by Honig and Bedi (2012) and Elliott et al. (2013) who draw attention to rather frequent explorations

of students' academic misconduct (including plagiarism), leaving practices on higher levels of academic hierarchies little reflected and explored. Hence, our conclusion from this rather simple analysis is that lack of findings on academics' plagiarism can be partially determined by the academic society's unwillingness to target the awareness of senior or established researchers by describing the issue of plagiarism in ethics codes.

Finally, considering a present discussions of the term *plagiarism* (e.g., Bouville, 2008; Clarke, 2009), the analysed Codes of Ethics provide just a very general characterisation which does not cover specific issues encompassed by the phenomenon. Thus, the Codes in Lithuanian academia cannot be regarded as an exhaustive instrument of education and prevention of plagiarism.

Results 2: Informal attitude

Previous analysis of the survey data revealed that there is lack of knowledge about existence of the Ethic Code in Lithuanian academic organizations among their members, and the extent of knowledge is related to the person's position in the organization (i.e. the higher the academic position the more chances that the person knows about the Code; the chances increase if the person in addition holds a managerial position in the administration) (Novelskaitė and Pučėtaite, 2012). However, an initial quantitative and qualitative review of the answers did not reveal any obvious differences in comments made by the respondents with or without administrative positions or between the answers of those who knew or did not know about existence of the Ethics Code in their organizations. For example, quantitative calculations demonstrate that respondents who had administrative positions at the time of the survey composed 22% of the study participants; their answers to the question under analysis compose 28% of all given answers; their answers compose 30% of those who mentioned the term *plagiarism* (cf: corresponding proportions for those who had no administrative positions are, respectively, 64%, 72% and 70%). Although it was obvious that thoughtfulness of answers positively relates to the academic position and that professors and associate professors are slightly more inclined to present their answers to the open-ended questions, the differences in quantitative terms are not significant (Table 2).

After the initial identification of the first unit of analysis, i.e. the answers involving references to plagiarism, 98 answers (i.e. 23% of 424 answers) were selected for further analysis. It is worth mentioning that an actual proportion of the answers drawing attention to plagiarism and related issues must be larger because (a) the total number of the answers (i.e. 424) also includes such void outgivings as "don't know", "have no opinion" or "cannot answer" and (b) in order to escape from equivocal interpretations, only direct references to plagiarism were included in the analysis meanwhile such outgivings as "there is no clear legal basis defining copyright issues" or "desperate cribbing, stealing others' ideas or even thoughts [are the main problems]" or "presenting ideas and results as one's own, without giving references to the actual authorship" and alike were not included.

In quantitative terms, it is interesting that most frequently plagiarism was mentioned in the answers of technical assistants, researchers and assoc. professors (33% and 31% of all answers provided by representatives of the groups). Slightly less frequently the

Table 2

Distribution of the study population by the main academic position and responding to the analysed question

Academic Position (the main)*	Survey population		<i>In your opinion, what are the most significant research ethics problems in Lithuanian scientific community and in the entire Lithuanian science system?</i>			
			Answers in total		Commentaries including notions of plagiarism	
	N	%	n	%	n	%
Professor	73	8	56	13	14	14
Principal Researcher	35	4	18	4	5	5
Assoc. Professor	156	17	91	22	28	29
Sr. Researcher	116	13	69	16	11	11
Lecturer	146	16	78	18	18	18
Researcher	27	3	13	3	4	4
Assistant	69	7	27	6	5	5
Jr. Researcher	78	9	33	8	6	6
Technical assistant	24	3	6	1	2	2
Other	59	7	33	8	5	5
N.i.	130	14	—	—	—	—
Total	913	100	424	100	98	100

* The answers below in this article are cited with information about the status of the respondent. Letter P stands for “Professor”; AP stands for “Associate professor”; L stands for “Lecturer”; JR stands for “Junior researcher”. If the respondent possessed an administrative position, abbreviation “admin.” is added to the information. Knowledge about existence of the Ethic Code in the organization is marked as Y_{EK} (“yes, the Code exists”), N_{EK} (“no, the Code does not exist”), DN_{EK} (“don’t know whether the organization has the Code”).

references were found in the answers of principal researchers, professors and lecturers (correspondingly, 28%, 25% and 23% of all answers in the groups). The least numerous references to plagiarism were found in answers provided by other (e.g. PhD students, administrators, engineers etc.) respondents and sr. researchers (correspondingly, 15% and 16% among all answers in the group).

Following the answers, plagiarism is “copying or re-narration of foreign and Lithuanian authors without referencing or giving inaccurate references to the sources” (AP, admin.; JR, DN_{EK}) or “presenting other authors’ results as one’s own” (AP, admin., N_{EK}), “plagiarism of results from literature and their presentation as one’s own results” (L, DN_{EK}). The third aspect of the definition (which, rephrasing Bouville (2008, p. 314), would be called “not real plagiarism”; see Clarke (2009) for a detailed discussion of the term) is provided in addition to copying of other authors with references and translations from other languages without citing by a professor who had no administrative position but knew about organizational Code of Ethics: “self-plagiarism is the largest problem in Lithuania—massive publication of the same studies in tens of publications, books, chapters, conferences etc.” Meanwhile other respondents insisted

on that “there is no clear definition of what plagiarism is” (AP, admin.). Moreover, “there is no developed system of plagiarism prevention; identified cases of plagiarism are not punished, which induces others to continue such activities” (L, Y_{EK}). Thus, “we do not have any legal basis for revoking degrees acquired by plagiarism” (P, DN_{EK}); in the absence of a legal mechanism “mere rhetorical, declarative provisions exist” (L, admin., DN_{EK}).

In several other answers plagiarism was related to students’ works as a problem at “master level and other works” (L), to which “attention is not paid” (L, Y_{EK}), and that “publicly known cases of plagiarism and falsification are not discussed in institutions, are not presented to students, to future doctoral students” (AP, admin, Y_{EK}).

However, in general, most frequently plagiarism was mentioned as the only or one of the most significant research ethics problems such as fabrication, bias, corruption etc. in Lithuania. Just in several answers plagiarism was denoted either as an “always arising question” or as “the most important, but not the most frequent problem”, which is “heavily controlled” or “tolerated” since “plagiarised works are accepted as original scientific studies”.

Summing up the presented findings, it can be concluded that in general almost 1/4 of the survey participants denoted plagiarism as one of the most important research ethics problems in Lithuania. Since opinions were expressed by respondents from different levels of an academic hierarchy, the findings indicate that the community is aware of the problem and perceives a need to change.

Conclusions

Analysis of Ethics Codes of 14 state and 3 non-state universities and 4 state research institutes in Lithuania revealed that a formal definition of plagiarism is not provided for all members of the community: some academic organizations still had no Codes in the middle of 2011; in some Codes plagiarism was not mentioned at all; the definition was not provided in all codes. Such results imply lack of a definition and clearness of the issue in Lithuanian academic community, which appears in answers of its individual members. Thus, the conclusion is simple: taking for granted that academia has deep awareness of the forms of plagiarism is deceptive and may not ensure self-regulation; clear formal rules are called for by the members.

Hence, a direction for further research is to estimate the need for the rules and development of corresponding initiatives. On the other hand, content analysis of the formal definitions of plagiarism and academic community’s attitudes to the phenomenon also reveals that, at least in particular cases, members of the community define plagiarism in broad terms (i.e. including translations and self-plagiarism) than organizational Ethic Codes. This finding presupposes a slightly opposite conclusion: members of the community possess the knowledge and the knowledge is not limited to formal definitions. Nonetheless, as rather frequent references to plagiarism as the most (or one of the most) important problems of research ethics in Lithuania suggest, the definition is not the main thing; other measures such as courses, open discussions on plagiarising and institutional procedures after detecting plagiarism are missing.

The results of analysis demonstrate that formal definitions of plagiarism appeal to two interrelated but different audiences: a rather anonymous academic community and students. Moreover, the definitions provided for relatively separate audiences are not the same and, following them, different groups of the academic community receive different amounts of information (and control). This is slightly echoed in several remarks by the community members who described plagiarism as one of the main research ethics problems in Lithuania. Possible outcomes of such a (relative, though) subdivision of the external locus of control, i.e. more stringent control of the community members who belong to the lowest stratum (i.e. students) and bypassing the highest levels of hierarchy are just speculative at the moment and need further empirical explorations.

Notwithstanding, it must be noted that the main shortage of the presented study is that the analysis of a formal definition and informal attitudes reveals only normative (which usually tend to be ideal and metaphysical) and subjective aspects (which is usually based on very personal experiences and feelings) of the issue. A factual situation as characterised by frequency of plagiarism, its forms and motivations etc. is not elucidated in this study. One of the reasons for this is that the data are taken from a research which originally focused on research ethics in general, which determined particularly formulated questions in the questionnaire survey and asking about plagiarism as one of potential ethical problems in research. Hence, identification of causes of academics' plagiarism in Lithuanian academic community lies in the future research field. Moreover, any justifiable claims that Lithuanian academic community as a post-soviet society faces many more cases of plagiarism than Western societies needs both a different comparative research and, on the other hand, deeper investigation into the interrelations between reported plagiarism and other, even ironic comments in open-ended questions of the survey, which can be done in the future as well.

References

- Bandura, Albert (2009/1986) *Socialiniai minties ir veiksmo pagrindai. Socialinė kognityvinė teorija.. [Social Foundations of Thought and Action. A Social Cognitive Theory]*. Vilnius: VU Specialiosios psichologijos laboratorija.
- Bouville, Mathieu (2008) Plagiarism: Words and Ideas. *Science & Engineering Ethics*, 14 (3), 311–322.
- Chester, Allan G. (1949) Richard Johnson's Golden Garland. *Modern Language Quarterly*, 10 (1), 61–67.
- Clarke, Roger (2009) Journal self-citation XIX: Self-plagiarism and self-citation—a practical guide based on underlying principles. *Communications of AIS*, 25 (19), 155–164.
- Cole, Sally and McCabe, Donald L. (1996) Issues in academic integrity. *New Directions for Student Services*, 73, 67–77.
- den Ouden, Hanny and van Wijk, Carel (2011) Plagiarism: punish or prevent? Some experiences with academic copycatting in the Netherlands. *Business Communication Quarterly*, 74 (2), 196–200.
- Elliott, Teresa L. and Marquis, Linda M., Neal, Catherine S. (2013) Business ethics perspectives: Faculty plagiarism and fraud. *Journal of Business Ethics*, 112 (1), 91–99.
- Elmore, Patricia B. (2010) Reporting standards for research publications. *Counseling Outcome Research and Evaluation*, 1, 19–29.

- Forster, Margaret and Loughran, Tim, McDonald, Bill (2009) Commonality in codes of ethics. *Journal of Business Ethics*, 90 (suppl. 2), 129–139.
- Furtado, Manuel L. (1950) Literary property: Common law protection of dramatic works. *California Law Review*, 38 (2), 332–340.
- Goodale, David (1938) Some of Walt Whitman's borrowings. *American Literature*, 10 (2), 202–213.
- Honig, Benson and Bedi, Akansha (2012) The fox in the hen house: A critical examination of plagiarism among members of the Academy of Management. *Academy of Management Learning & Education*, 11(1), 101–123.
- Hudson, Richard (2008) Researchethics and integrity—the case for a holistic approach. *Research Ethics*, 4, 131–140.
- Insley, Robert (2011) Managing plagiarism: a preventative approach. *Business Communication Quarterly*, 74 (2), 183–187.
- Ivanauskas, Vilius (2006) Sovietinių biurokratų darbo etika, neformalios rutinos ir planavimo sistemos trūkumai (Lietuvos atvejis) [Work ethics of soviet bureaucrats, informal routines and shortages of the planning system: the case of Soviet Lithuania]. *Filosofija. Sociologija*, 4, 1–12.
- Jameson, Daphne A. (1993) The ethics of plagiarism: how genre affects writers' use of source materials. *Business Communication Quarterly*, 56, 18–28.
- Kaptein, Muel and Schwartz, Mark S. (2008) The effectiveness of business codes: A critical examination of existing studies and the development of an integrated research model. *Journal of Business Ethics*, 77 (2), 111–127.
- Liles, Jeffrey A. and Rozalski, Michael E. (2004). It's a matter of style: A style manual workshop for preventing plagiarism. *College & Undergraduate Libraries*, 11 (2), 91–101.
- Martin, Brian (1994) Plagiarism: A misplaced emphasis. *Journal of Information Ethics*, 3 (2), 36–47. Available from: <http://www.bmartin.cc/pubs/94jie.html> [Accessed on January 29, 2013].
- Martin, Daniel E. (2012) Culture and unethical conduct: Understanding the impact of individualism and collectivism on actual plagiarism. *Management Learning*, 43(3), 261–273.
- Martin, Daniel E. and Rao, Asha, Sloan, Lloyd R. (2009) Plagiarism, Integrity, and Workplace Deviance: A Criterion Study. *Ethics & Behavior*, 19 (1), 36–50.
- Masterson, Salathiel C. (1940) Comment: Copyright: History and Development. *Law Review*, 28 (5), 620–632.
- Maurer, Hermann, Kappe, Frank and Zaka, Bilal (2006) Plagiarism—A Survey. *Journal of Universal Computer Science*, 12 (8), 1050–1084.
- McCabe, Donald L. and Trevino, Linda K. (1993) Academic dishonesty: honor codes and other contextual influences. *Journal of Higher Education*, 64 (5), 522–538.
- McCabe, Donald L. and Trevino, Linda Klebe (1993) Academic Dishonesty: Honor Codes and Other Contextual Influences. *Journal of Higher Education*, 64 (5), 522–538.
- Novelskaitė, Aurelija and Pučėtaitė, Raminta (2011) Mokslinių tyrimų etikos būklė Lietuvoje universitetų ir mokslinių tyrimų institutų darbuotojų požiūriu. Kaip pakeisti esamą situaciją? [Attitude of universities' and research institutes' staff towards status of research ethics in Lithuania. How to change current situation?] In: Norvaiša, R. (ed.) *Mokslas ir etika*. Vilnius: LMS, 29–62.
- Novelskaitė, Aurelija and Pučėtaitė, Raminta (2012) Mokslinių tyrimų etikos būklė Lietuvoje: ką mano Lietuvos mokslo bendruomenė? Pirminiai nacionalinės apklausos rezultatai. [Research ethics status in Lithuania: what does Lithuanian scientific community think? Preliminary results of national survey] *Socialinių mokslų studijos*, 4 (4), 1371–1395.

- Novelskaitė, Aurelija, Pučėtaitė, Raminta and Lipnevič, Anna (2011) Defining Research Ethics Legally: Liaison between International and National Documents. The Case of Researcher's Safety as a Research Ethics. *The International Journal of Science in Society*, 2 (3), 219–234.
- Parker, William R. (1945) Winstanley's lives: An appraisal. *Modern Language Quarterly*, 6 (3), 313–318.
- Pučėtaitė, Raminta and Lämsä, Anna-Maija (2008) Developing organizational trust through advancement of employees' work ethics in a post-socialist context. *Journal of Business Ethics*, 82, 325–337.
- Ryan, Leo V. (2006) Current ethical issues in Polish HRM. *Journal of Business Ethics*, 66, 273–290.
- Sisti, Dominic A. (2007) How Do High School Students Justify Internet Plagiarism? *Ethics & Behavior*, 17 (3), 215–231.
- Vasiljevienė, Nijolė and Freitakienė, Rasma (2002) Are we ready to accept European standards? In Vasiljevienė, N. and Jeurissen, R.J.M. (eds.), *Business Ethics: From Theory to Practice* [Vilnius: Vilnius University] (pp. 172–195).
- Žiliukaitė, Rūta, Ramonaitė, Ainė, Nevinskaitė, Laima et al. (2006) *Neatrasta galia: Lietuvos pilietinės visuomenės žemėlapis* [Undiscovered power: map of the civil society in Lithuania]. Vilnius: Versus Aureus.

Acknowledgements

Dr. Aurelija Novelskaitė thanks Advanced Certificate in Research Ethics for Central and Eastern Europe Program organized by Union Graduate College of Mount Sinai School of Medicine (Schenectady, NY, USA) and Vilnius University (Lithuania) for partial funding of the accomplishment of the study. Special thanks go to the professors M. Strosberg, E. Gefenas, S. Philpott and others who strongly encouraged and supported A. Novelskaite's works in the field of research ethics.

Authors

Aurelija Novelskaitė, Novelskaite@ktl.mii.lt, Institute of Sociology, Lithuanian Social Research Center, Lithuania

Raminta Pučėtaitė, Raminta.Pucetaite@khf.vu.lt, Kaunas Faculty of Humanities, Vilnius University, Lithuania

Copyright © 2013 A. Novelskaitė, R. Pučėtaitė: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

PAPERS—SECTION IV
INNOVATION IN USE OF E-TOOLS AND
TECHNOLOGIES FOR ADDRESSING PLAGIARISM

PLAGIARISM DETECTION IN SLOVAK TEXTS ON THE WEB

Daniela Chudá, Jozef Lačný, Maroš Maršalek, Pavel Michalko, Ján Súkeník

Abstract: Nowadays, access to all kinds of information is greatly simplified. The Web contains enormous amounts of documents available for free. One negative consequence of this is the increase of plagiarism, not only in the fields of research and education, but also in journalism etc. Authors incline to plagiarism for various reasons and the Internet makes it extremely easy. A lot of effort and research have been devoted into developing tools that detect plagiarism on the Web. While the majority of these tools worked correctly during our tests (some had better results than the others), we have found room for improvements. First of all, they are designed to work with English language and therefore their accuracy for texts in more complicated languages (Slovak for example) is not satisfactory. Moreover, the presentation of the results to the user was often very unintuitive and limited.

We have built an application which provides automated plagiarism detection in Slovak texts among the documents on the Web. Our tool utilizes reliable search engines like Google and Bing. The main advantage of this approach is that it does not require building a custom index of the Web which would be time and resource consuming. Additionally, we provide advanced visualization of results and we also offer an option to manually tweak them. The application is available at <http://wanted.fiit.stuba.sk>.

Introduction

Problems associated with plagiarism started to rise quickly with the development and accessibility of the Internet. Improvements of internet search engines and their widespread popularity influenced the way of retrieving and utilising information sources. Most people nowadays prefer digital information sources available on the Internet.

Why is plagiarism considered to be a problem? We can get our answer by looking at the field of education. Students are the ones who mostly incline towards using the Internet as primary source of information for their work and who often fail to reference their sources properly. Sometimes the reason might just be carelessness or lack of experience. However, the quality of knowledge gained by students who intentionally commit plagiarism is often decreased (Chudá et al., 2012).

Another example is the field of research. When working on a research project, one should rely mostly on his or her own knowledge and experience. However, it is usually helpful or even necessary to use related work from other authors. In these cases it is required to clearly identify all the references so that author's own contribution can be clearly identified. Plagiarism can also be a problem in many other fields, for example in mass-media (stealing content from each other).

To alleviate negative impacts of plagiarism it is necessary to be able to discover it. That means comparing all the new works to every paper, book or article on similar or overlapping topic available not only in the libraries but also on the Internet. When trying to detect plagiarism manually, the effort required is simply too big and can even

be multiplied if the author tried to disguise it. It is therefore desirable to fully automate the process of measuring similarity between documents.

We have designed and developed a tool that deals with this issue. Our application provides automatic similarity evaluations of submitted document against related documents available on the Internet. We have focused on processing texts in Slovak language, since there are almost no tools optimised for languages other than English. Our main goals are improving quality in education, boosting creativity of students and motivating journalists to create original content. Indirectly, our tool can provide help when trying to find related information sources. We faced several problems in order to create the application and our approach is presented in this paper.

In the paper, we firstly take a quick look at the existing research and tools. We continue with the high-level overview of our method and consequently describe individual steps of the process in greater detail. At the end we outline our experiments and discuss possible use cases of our application.

Related work

We divide the problem of plagiarism detection in Slovak language into multiple smaller problems.

First of all it is necessary to extract raw text from different document formats. Some of them contain a lot of formatting information (DOC, DOCX), others are designed for printing (PDF) which means that the formatting information even takes precedence over the text information and its semantics (titles, captions, etc.). Extraction of raw text from HTML documents is even more complicated because of all the clutter that websites contain (navigation, headers and footers, advertisements, etc.). There has been done a lot of research in this area (Nguyen Minh Trung et al., 2011), (Weninger, 2010) and many tools were implemented (The Apache Software Foundation, 2013).

To make text processing easier and results more accurate, the input document should be divided into smaller parts. Multiple ways of achieving this are outlined in (Pataki, 2003).

Another part of text pre-processing especially important for Slovak language is the lemmatization. Slovak Academy of Sciences provides morphological database which contains 2.5 million words (approximately 75000 different lemmas) and can be used for this purpose.

Very important problem is building of search queries. These queries must reflect the source text so that the same or similar texts will be found during the web-search. This is the crucial step when trying to find related documents on the Internet (Butakov and Shcherbinin, 2009).

The core of plagiarism detection is comparison of texts or documents. Lots of approaches for performing this task were suggested, each with its own advantages and disadvantages (Genčí et al., 2009). As examples we choose methods based on n-grams (Češka, 2007), longest common substrings (Chong et al., 2010) or stylometry (Clough, 2000).

The research in the field of plagiarism detection has resulted in a big number of plagiarism detection tools. We divide them into two groups: (1) the tools that use

Table 1

Comparison of some of the tools that detect plagiarism on the Web

	Turnitin	EVE2	DOC Cop	PlagScan	Copyscape
Free	no	no	yes	no	subset/no
Detection of citations	yes	no	no	no	no
Advanced support for Slovak language	no	no	no	no	no
Detection of anti-plagiarism	yes	no	no	no	no
Support for local corpus	yes	no	limited	yes	no
Web-search	custom	plugins	Bing	Yahoo!	Google, Yahoo!
Another limitations	known methods of cheating	platform dependency	input text limit	lower effectiveness	designed for websites

only local corpus of documents, (2) the tools that use more general, shall we even say global corpus—the Web. An example from the first group is CRZP (central repository of final theses). CRZP is used in Slovak Republic to ensure that the theses (bachelor, diploma, etc.) do not contain plagiarism. CRZP only detects plagiarism in local corpus of documents primarily made of all the theses ever submitted to it.

One of the tools that are capable of detecting a plagiarism on the Web is Turnitin (iParadigms, LLC., 2013). Main advantages of Turnitin are extensive local corpus of documents which includes documents from the Web, detailed results presentation and visualization, support for quotations of external sources, constant development and improvements. The biggest limitations of Turnitin are lack of advanced support for Slovak language and publicly available methods of cheating.

Other similar tools are: EVE2 (Essay Verification Engine) (CaNexus, 2000), DOC Cop (McCrohon, 2013), PlagScan (Resonet UG, 2013) or Copyscape (Indigo Stream Technologies, 2013). We summarize comparison of all mentioned tools in table 1. These tools and many others are compared more extensively in (Bull, 2001) or (Plagiats Portal, 2012).

All of the tools mentioned above share one common disadvantage. They can process texts in different languages, but (to our knowledge) none of them considers specifics of those languages. This is particularly important when dealing with partially modified copied text. Using stop-words removal, performing lemmatization etc. can yield more accurate results and more precise plagiarism detection.

Method description

In this chapter we present our own solution for plagiarism detection on the Web. We have designed a unique process that deals with this problem and one of most important characteristics of our solution is the utilization of already available Internet search engines like Google and Bing.

As we have mentioned before, the problem we are dealing includes many sub-problems which had to be solved. Therefore we have split this process into several steps.

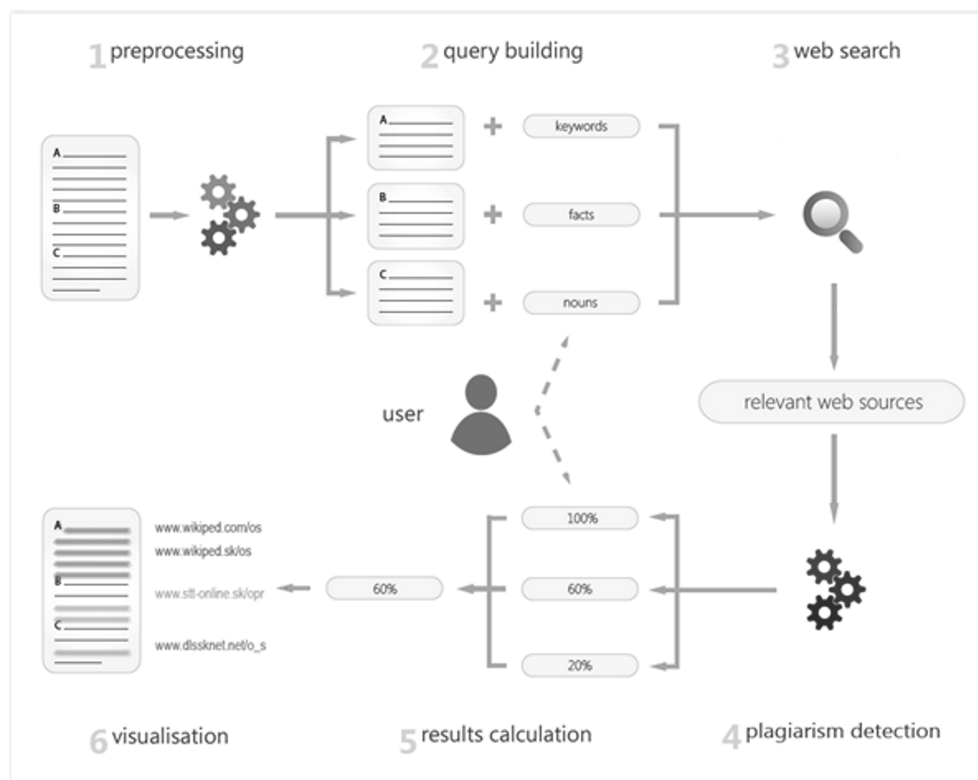


Figure 11.2. Our solution for the process of plagiarism detection on the Internet

For each of these steps we have designed an approach which enables us to achieve best performance and most accurate results.

The six steps of the whole plagiarism detection process (see figure 1):

1. pre-processing and chunking of input document,
2. construction of search queries,
3. web-search for related documents or possible sources of plagiarism,
4. plagiarism detection,
5. calculation of similarity level for the entire document,
6. visualization of the results.

These steps are described in detail in the following sections.

Text pre-processing

After the user submits a document to our application, we perform several steps of pre-processing before the analysis can be started. Firstly, we have to extract raw text from the document. We have experimented with multiple tools or libraries and at the end

we have settled with Apache Tika for formats such as PDF, DOC, DOCX or pure text and Readability for HTML documents.

In the next step we split the raw text into smaller parts—chunks. Processing of these chunks is more suitable for plagiarism detection, since only few sentences are usually copied. Moreover, different parts of text can be copied from different sources.

Query building

The first major issue we needed to solve was how can we possibly compare a submitted document with every document on the Web? This is clearly not possible. In fact we can only compare input document with very small subset of all the documents available on the Web. So the question is immediately changed to how can we obtain this subset of documents so that it will (with reasonably high probability) contain the documents we are searching for? The documents in this subset will probably have the same topic as the input document, which means they will contain similar keywords. This means that we need a full-text index of the Web and some search engine built on top of this index.

From this emerged the next question: how should we use such search engine? If we obtain keywords from the whole input document and use those as the search query, the probability of finding correct results (documents used as the source of plagiarism) is rather low. In fact, the longer the input document, the lower is the probability. On the other hand, we can use every sentence as the search query, or more general, every n-gram of the input document. This way it would be possible to detect plagiarism almost always. The number of web-searches would however be very high, the subset of documents to examine would be very big and the whole process would take a lot of time. Therefore we needed to find a solution between these two extremes.

In order to design a good query building algorithm, we needed to know how search engines work. We have done experiments with Google and Bing. Results of one experiment are summarized in table 2. Since we focus on documents in Slovak language, we decided not to translate sample queries.

We have obtained few very important conclusions from our experiments:

- query does not need to contain consecutive words,
- order of the words in the query does not matter at all,
- query does not need to contain many words, few are usually enough,
- query should not contain any words that are not present in the original document,
- query should not contain any words that were partially changed (prefixes, postfixes or any other modifications of words as a result of inflection or conjugation),
- if correct result is found, it is always in top results (usually first, but we have seen second or third).

Results of our experiments suggest that the query can be built from document's keywords, but with two big and important constraints. Selected keywords have to be exactly the same in the original document as in the plagiarized document, no modifications are allowed at all. That is rather unfortunate when considering that we

Table 2

Manual experiments performed to find best automatic query building method

Description	Search query	Result count		Position of the original article in the results	
		Bing	Google	Bing	Google
– initial query	čierne diery považované nutné pozerat' hľadiska mimoriadne dôležitú úlohu vzniku galaxií	3	132	1	1
– as no. 1 – changed “považované” → “považovali” – changed “nutné” → “treba” – changed “dôležitú” → “podstatnú”	čierne diery považovali treba pozerat' hľadiska mimoriadne podstatnú úlohu vzniku galaxií	1	113	Θ	Θ
– as no. 2 – removed word “treba”	čierne diery považovali pozerat' hľadiska mimoriadne podstatnú úlohu vzniku galaxií	1	113	Θ	Θ
– as no. 3 – removed word “podstatnú”	čierne diery považovali pozerat' hľadiska mimoriadne úlohu vzniku galaxií	8	125	Θ	Θ
– as no. 4 – removed word “považovali” – the query is equal to no. 1 with three words removed	čierne diery pozerat' hľadiska mimoriadne úlohu vzniku galaxií	102	128	1	1
– as no. 5 – changed “pozerat'” → “pozrieme”	čierne diery pozrieme hľadiska mimoriadne úlohu vzniku galaxií	8	1080	Θ	Θ
– as no. 6 – changed order of words	mimoriadne úlohu vzniku galaxií čierne diery pozerat' hľadiska	96	1280	1	1
– short query, removed majority of words	čierne diery vesmírne pasce	15	6400	1	1
– as no. 8 – added word “astronómia” (part of original article)	čierne diery vesmírne pasce astronómia	141	101	1	1
– as no. 9 – added word “pohľadu” (not part of original article)	čierne diery vesmírne pasce astronómia pohľadu	73	4150	Θ	2
– another query – duplicated word “stiahnuť”	považované pasce stiahnuť všetko nebezpečne astronómia čierne nutné úplne mimoriadne	1	7	1	Θ

deal with an inflectional language. Secondly, if only as little as one word in a query is not present in the original document, search engine will probably not find it.

As we stated earlier, we divide the input document into small chunks. For each chunk we build one search query. Query contains few words from chunk (less than 10) which have the high probability of satisfying the two constraints mentioned above. This query

is the core of our method and allows us to find documents which contain similar or exactly the same text as the text in the chunk.

We have designed multiple methods for query building, such as:

- facts extraction—numbers, dates, names, in other words parts of text that are almost impossible to modify without changing the meaning of the text,
- nouns selection—although not impossible, nouns can also be difficult to modify,
- most frequent words—simple keywords of individual chunks,
- TF-IDF based selection of keywords—regular TF-IDF uses whole documents and returns keywords for one document; we use chunks and we are interested in keywords for one chunk,
- random selection,
- combinations of the above.

We performed several tests to determine the best method (see table 3). The primary method which we currently use is combination of facts, nouns and most frequent words.

Web-search

Next step of the process is the web-search. We do not have enough resources to create our own index of the Web and search engine. It is incredibly complex topic and not our focus at all. Moreover, there are companies, organizations or other initiatives that already have web-search engines running, so we decided to outsource this part of the process. The two major players in this area are currently Google and Microsoft with their Bing. We decided to support both of them. Currently we do not utilize digital libraries (e.g. ACM, IEEE) because they do not provide free access. We rely solely on the results from web-search engines.

At this point of the overall process we have the input document divided into chunks and search query built from each chunk. Our application now starts performing web-searches using all the available queries (that means one web-search per chunk). The third conclusion derived from the experiments described in section about query building is that it is sufficient to evaluate plagiarism in only few top results. Therefore when we receive results of the web-search, we only download first 5 documents.

Similarity detection

So far we have divided input document into chunks, created search queries and downloaded search results for each available chunk. During this step we pre-process documents acquired from the Internet (just like the input document but with the exception of chunking). We also optionally perform lemmatization on both (1) the text of the chunk as well as (2) the text of the document from the Web.

After this additional processing we calculate similarity by comparing the n-grams (on the word level) of the chunk with the n-grams of search result. The outcomes of this comparison are (1) the levels of similarity between a chunk and the respective web document and (2) the list of identical n-grams.

Similarity level calculation

In this step we merge partial results obtained in the previous step. The result of this process is one number that represents final level of plagiarism of the whole input document. During this process we had to solve some typical problems, for example the same text being found in multiple web documents, or one chunk being copied from two non-overlapping sources.

Visualization

Our main contribution lies in the first five parts of document processing, nevertheless we consider proper and user friendly visualization of results to be very important for this kind of application. We have built the user interface that tries to preserve balance between user-friendliness, ease-of-use and proper visualization of all important results of document analysis.

One of the problems here was correct highlighting of copied parts of the input document. As we mentioned earlier, the documents are being modified during the analysis (removing whitespace, special characters, stop words or lemmatization). Because of these modifications it is difficult to correctly highlight copied parts of the original (not modified) text for the purpose of presentation to the user. The documents may moreover sometimes contain parts of copied text that are in fact correctly cited or false positives. For these cases our application offers an option to mark every detected positive text matching as correctly cited. After this operation, the selected text is excluded from plagiarism evaluation and results are recalculated and refreshed in real-time so that the user can always see the most up-to-date information.

Architecture and technology

As we mentioned earlier, we have developed an application that implements described solution. The architecture of the application conforms to standard client-server model with business and data layers located on a centralised server and presentation layer running in user's web browser as a thin client.

The business layer is implemented in Ruby on Rails and provides the core functionality of our solution through RESTful web services. This implementation allows clear separation between business and presentation layer and the functionality of our solution can be used by multiple different clients.

To ensure the best possible user experience we have chosen HTML5 to be our main technology for presentation layer development. Our goal was to provide intuitive and easy to use graphical interface.

Experiments

The query building process plays the key role in our solution. We have spent a lot of time designing query building methods as well as analysing their performance. In one experiment we analysed plagiarism level of a set of artificially prepared documents using different query building methods and search engines. These documents contained

Table 3

Comparison of query building methods using different search engines

	random		facts + nouns + most frequent words		facts + nouns + most freq. words + random		facts + nouns + tfidf + typos		expected
	Google	Bing	Google	Bing	Google	Bing	Google	Bing	
1.	87%	30%	87%	49%	79%	38%	62%	22%	87%
2.	95%	79%	100%	84%	97%	84%	95%	84%	100%
3.	45%	27%	37%	26%	21%	27%	22%	21%	47%
4.	21%	1%	42%	0.01%	51%	0%	36%	0.01%	70%

text copied from one or more sources available on the Internet. This experiment shows that our query building methods are in fact able to find correct documents.

Table 3 shows the results of the experiment. There are the three most significant query building methods compared to the random query building method as well as the expected results.

The expected result is the real plagiarism level of an artificially prepared plagiarized document. The results show that the best query building method for finding relevant documents on the Internet is combination of facts, nouns and most frequent words. We also found that search for similar documents using Google is much preferable compared to Bing when using queries in Slovak language.

Conclusions

The result of our work is a tool whose primary goal is to help with plagiarism detection. We took advantage of current research and performed necessary experiments in order to make this tool efficient. Thanks to our solution the whole process of document analysis is performed only by just a few simple user interactions.

Comparison to domain expert

The human judges are very good at determining whether some particular part of the text is copied from some specific source. But as soon as they try to find those sources themselves, various problems come up. First of all, such plagiarism evaluations may be different each time. Manual searches and manual comparisons can also be very slow and inaccurate, since the number of possible sources (even about one specific topic) is usually big and one person cannot possibly be familiar with all of them. Walenstein et al. (Walenstein, et al., 2003) addressed similar issue in their work with source code clones. They conclude that single judge cannot be trusted to give unbiased answers (they actually used multiple judges who always had to go through consensus-building process). Although Walenstein et al. focused on different problem, we believe that their conclusion can be generalized. A domain expert's opinion can therefore be considered helpful at best, not obligatory.

Secondly, we have never intended our tool to replace the domain experts or to compete with them. We believe that no matter how good the automatic plagiarism detection is, it will never be possible to completely leave out the final judgement of the domain expert. Our goal is just to simplify and automate the repetitive tasks which are part of the plagiarism detection process. We hope that it will make the whole process not only easier and faster, but also more objective and accurate. Purpose of our tool is therefore to support the domain experts and to minimize the impact of all the disadvantages of manual plagiarism detection.

Use cases of an application

We have identified several use cases for our application. The main use case is of course plagiarism detection where a typical scenario would be a teacher checking assignments of his or her students. Besides plagiarism detection, we think it would also be helpful if students or researchers could check their papers before submitting them and therefore had a chance to correct the references to other sources in case of any mistakes.

Another interesting use case is effort analysis. If we look at the elementary or high schools, it is not unusual that the teachers want the students to write papers which consist exclusively from the work of other authors. In such cases it is difficult and time consuming to find out whether a student copied the text from one or two sources or if the student found multiple interesting sources and combined them in a creative way. The teacher could use our application to find the list of sources used in such papers and consider this during paper evaluation.

In its core, our application tries to find documents which address the same topic as the input document. This feature is just one part of our plagiarism detection solution, but it could be utilised on its own. The goal would be to discover new sources of information on certain topic that were yet unknown to the user. Such feature could for example find additional literature just by submitting an unfinished paper. Despite the fact that our method is capable of providing this feature, our user interface, in its current version, is not adjusted for it. We are however planning to update the application because we think that the ability to conduct literature review on a topic would be useful.

Other languages

This application was developed mainly for processing texts written in Slovak language, but it can be used with different languages too at the cost of slightly lower precision. On the other hand, if the language dependent parts of application will be replaced (e.g. normalisation or lemmatization modules) the final results would be as precise as in case of the Slovak texts.

Our application is available at <http://wanted.fiit.stuba.sk> and we encourage everybody to try it out.

References

- Bull, J., Collins, C. et al. (2001) *Technical review of plagiarism detection software report*. University of Luton. JISC publication
- Butakov, S. & Shcherbinin, V. (2009) On the Number of Search Queries Required for Internet Plagiarism Detection. *Ninth IEEE International Conference on Advanced Learning Technologies*, 482–483
- CaNexus (2000) EVE2. Available from: <http://www.canexus.com/> [accessed 12 January 2013]
- Češka, Z. (2007). Využití N-gramů pro odhalování plagiátů. In *Proc. of Konferencia o informačných (inteligentných) technológiách – aplikácie a teória 2007*, Slovakia
- Chudá, D., Návrat, P., Kováčová, B. & Humay, P. (2012) The Issue of (Software) Plagiarism: A Students' View. *IEEE Transactions on Education*, 55 (1), ISSN 0018-9359, 22–28
- Chong, M., Specia, L. & Mitkov, R. (2010) Using Natural Language Processing for Automatic Detection of Plagiarism. In *Proc. of the 4th International Plagiarism Conference (IPC-2010)*
- Clough, P. (2000). Plagiarism in natural and programming languages: an overview of current tools and technologies Keywords. 1–31
- Genčí, J., Maťašovská, M. et al. (2009) Plagiarism and ways of its restriction. *UNINFOS 2009*, 25–27 November, Nitra, Slovakia
- Indigo Stream Technologies. (2013) Copyscape. Available from: <http://www.copyscape.com/> [accessed 10 April 2013]
- iParadigms, LLC. (2013) Turnitin. Available from: <http://turnitin.com> [accessed 12 January 2013]
- McCrohon, M. (2013) DOC Cop. Available from: <http://www.doccop.com/index.html?nc=35190108> [accessed 12 January 2013]
- Nguyen Minh Trung, Nguyen Duc Tam & Nguyen Hong Phuong. (2011) Using main content extraction to improve performance of Vietnamese web page classification. In *Proc. of the Second Symposium on Information and Communication Technology (SoICT '11)*. ACM, New York, USA
- Pataki, M. (2003) Plagiarism detection and document chunking methods. Available from: http://dsd.sztaki.hu/people/mate_pataki/200305_WWW2003.pdf [accessed 31 January 2013]
- Plagiats Portal. (2012) Collusion Test 2012—Collusion Detection System Test Report 2012. Available from: <http://plagiat.htw-berlin.de/software-en/2010-2/> [accessed 10 April 2013]
- Resonet UG. (2013) PlagScan. Available from: <http://www.plagscan.com/> [accessed 10 April 2013]
- The Apache Software Foundation (2013) Apache Tika. Available from: <http://tika.apache.org/> [accessed 12 January 2013]
- Tim Weninger, William H. Hsu & Jiawei Han (2010) CETR: content extraction via tag ratios. In *Proc. of the 19th international conference on World wide web (WWW '10)*. ACM, New York, USA, 971–980
- Walenstein, A. et al. (2003) Problems Creating Task-relevant Clone Detection Reference Data. In *Proc. of the 10th Working Conference on Reverse Engineering*, 285–294

Acknowledgement

This work was partially supported by the Scientific Grant Agency of the Slovak Republic, grant No. VG1/0971/11 and is the partial result of the Research & Development Operational Prog. for the project ITMS 26240220039, co-funded by the ERDF.

Authors

Daniela Chudá, chuda@fiit.stuba.sk,

Jozef Lačný, lacny.jozef@gmail.com,

Maroš Maršalek, maros.mars@gmail.com,

Pavel Michalko, michalko.pavel@gmail.com,

Ján Súkeník, sukenik08@student.fiit.stuba.sk

Slovak University of Technology, Faculty of Informatics and Information Technologies,
Slovakia

Copyright © 2013 Chudá, D., Lačný, J., Maršalek, M., Michalko, P., Súkeník, J.: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

IMPROVING PLAGIARISM DETECTION THROUGH THE DOCUMENT CONTENT SEGMENTATION

Tomáš Kučečka, Daniela Chudá

Abstract: Topic extraction from text documents is an important object of research these years. It can be used, for instance, in an information retrieval to group similar results together returned by a web browser or to simplify plagiarism detection in a given document corpus. In this paper we propose a novel approach to document segmentation that should be helpful with plagiarism detection. It is based on analysing keyword positions in a text document. Based on these analyses we try to segment document's content and identify similarity between documents based on the similarity between found segments. Therefore, this approach shows how much about a document can be told from the distribution of its keywords. We think that through this approach we are able to better identify relationships between text documents (between their segments) and come to the more accurate plagiarism identification in a given document corpus. We believe that two documents with similar segments should gain more attention in the plagiarism detection process. In this article we present our first experiments with this approach that we carried out on the student's assignments. We used results from an existing plagiarism detection system PlaDeS to evaluate our approach.

Introduction

Plagiarism is becoming a serious problem these days especially in academic environment. Academic institutions have to take actions to prevent and detect such behaviour in order to keep their credibility. This has become much more difficult with the availability of online resources, which number rapidly increases every year. This makes it a lot easier for people to plagiarise.

Although many people think of plagiarism as rewriting someone else's work without giving the credit to the original author, this is not its only meaning. According to the plagiarism.org resource, the term "*plagiarism*" can be defined as "to steal and pass off (the ideas or words of another) as one's own" or "to use (another's production) without crediting the source".

For the two main problems that are associated with plagiarism detection we consider source identification and paraphrasing. The source identification problem has become much more serious with the spread of Internet and availability of online resources. Without having a source document from which the plagiary comes, it is very difficult to detect a plagiarism. Paraphrasing is another main problem that can be done on different levels. The higher the paraphrasing level, the more difficult is its identification. In this paper we focus on the second problem—paraphrasing identification.

We propose a novel approach to plagiarism detection among texts written in natural language. This approach is based on document segmentation and is mainly intended for longer documents, e.g. bachelor or diploma thesis. The idea of the solution is to model a text document as a mixture of segments, where each segment reflects a topic on different granularity. A document can contain several of these topics with a minimum of one. Through the segments found in documents, we propose document candidates

that should be checked on plagiarism in other way than using standard approaches. For instance, a 3-gram approach can be used to check plagiarism between documents that are less related based on the found segments. On the other hand, documents that contain higher number of similar segments can be checked using more effective comparison techniques with higher computational complexity.

Our approach distinguishes two types of segments—global and local. Global segment covers a global topic of a document, i.e. the topic of a whole document. It holds that every document has exactly one global segment. Local segments represent a subtopic of a document and their number in a document is arbitrary starting from zero. Based on this, every document contains at least one global segment.

Related work

As our approach deals with plagiarism identification based on document topic detection through document segments, we focus here on related work from this area of research.

Finding relationships between text documents based on their topic has gained a significant importance in last years. The existing approaches in this domain vary from different clustering techniques based on frequent-itemsets to probabilistic approaches. Better results in similarity detection can be achieved by single or multi-document segmentation that try to identify where the topic of a text document changes.

The main drawback of the existing approaches is that they frequently consider document as a single topic. This is, in our opinion, correct for shorter texts but becomes a problem for longer documents. A typical example of a text document with several discussed topics is diploma thesis, which is basically structured on the analysis, design, experimental and technical part. For each of these parts usually a different set of keywords is typical. Our approach models a document such way, that it identifies these different parts and their keywords. Basically, one can consider our approach as a single document segmentation method. This is partially true but our approach also extracts keywords that characterize the document as a whole. We call these keywords global keywords.

When designing our own approach, we were inspired by the existing pLSA (probabilistic Latent Semantic Analyses) approach presented by Hofmann, T. (1999) and LDA (Latent Dirichlet Allocation) approach presented by Blei, M.D. et al. (2003). The LDA is basically an enhancement of pLSA. Both of these approaches assign each document with a number of topics, each with different probability based on the words the documents contain. The background behind these approaches is the mixture model, which models the whole document corpus as a distribution of different topics represented by words from a given corpus. Both pLSA and LDA require that a user must know a number of topics to which documents from corpus are going to be assigned. Our approach has no such restriction.

Ambwani, G. et al. (2010) proposed a novel approach to topic segmentation. The authors hold to the idea that the used vocabulary reflects the topic continuity. A text document using this approach is modelled as a graph of RI's (Relevance Interval). RI represents a range of term influence (basically it is a range where the term occurs at

most). Based on the RI's in the constructed graph (the way that the RI's are connected) a document is segmented. Other work on the topic of document segmentation was presented by Sun, B. et al. (2007). They introduce MI (Mutual information) and WMI (Weighted Mutual Information) indexes for mutual information computation. This approach belongs to the multi-document segmentation. The MI index measures how distinct the two segments are, which is a different approach compared to the standard approaches based on the similarity calculation (e.g. cosine similarity). The WMI index is simply a weighted version of MI computation that should perform better.

Proposed approach

In this section we describe a novel approach to finding relationships between text documents based on their topic. The overall design emerged from our previous work presented in Kueka, T. (2012) and Kueka, T. et al. (2012). We assume only documents written in natural language. Basically, we can divide the whole approach to three main phases:

- Preprocessing
- Segmentation
- Relationships identification

In the preprocessing phase of a document we used standard techniques such as stop words removal, lemmatisation and synonyms replacement. In segmentation phase we identify document segments that represent the document's topics and subtopics. Then, in the third phase we identify relations between segments based on the keywords these segments share.

Segmentation process

The segmentation process is based on calculation of keyword distribution vectors over text blocks, where text block represents a string of a 260 letters. This length was estimated empirically. Text blocks are extracted from document's content. After the extraction process is finished, the following steps are applied:

1. *Keyword extraction.* Based on the tf-idf value a set of keywords is extracted. For every document words with higher weight than 0.03 are selected. Notice, that this threshold is quite low and the number of selected words for a document can therefore be very high. Our aim is not to select as few words as possible but rather to filter out only common words. From now on, we refer to all selected words in this step as keywords.
2. *Finding keyword distribution vector.* The frequencies of keywords (tf values) over extracted text blocks are first calculated. This calculation is done per document. By dividing the keyword frequencies over text blocks (TB) with the total tf value of a keyword in a document, we get a distribution vector. This vector represents a distribution of a keyword over document TBs. In the left picture (a) of Figure 1 an example of a distribution vector of a keyword w is shown. The right picture (b) shows two similar distributions of two different words.

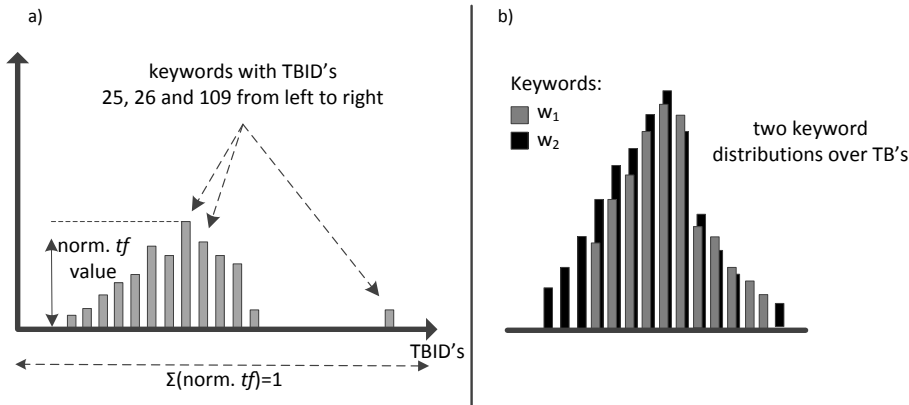


Figure 11.2. a) An example of a keyword distribution; b) Shows two similar distributions of two different keywords

By TBID we mean the ID of a text block. Every text block has an ID that is unique per document. This ID represents the order of a text block in a document. The first extracted TB has ID 0.

3. *Comparison of keyword distributions.* We compare the keyword distributions per document using the Jensen-Shannon (JS) divergence presented in Cedeño, B.A. et al. (2009) to find out if two keywords occur in a text document in similar text blocks. To such keywords we refer as similar. The output of JS distance is a value that represents the amount of work that has to be done to transform one distribution vector on another. This value lies in the $[0, 1]$ interval.
4. *Segment building.* Keywords with similar distribution form a segment, i.e. every segment is represented by a set of similar keywords. In order to tell if a keyword belongs to a local or a global segment, we calculate the variance for all distribution vectors. Higher variance indicates that the keyword is more sparsely present in a document's content. In such case we consider the keyword global. The distributions with highest variance form a global segment of a document. Other keywords are local and therefore can form local segments. Two different distributions of a keyword, one of which is local and one of which is global, are depicted in Figure 2.

A minimum number of keywords in local segment is 3. This value was estimated empirically based on the experiment results. Otherwise the segment will not be created. All local keywords that do not belong to any segment are dropped.

Relationships identification

First, an undirected segment graph G is constructed from found segments. Vertices of G represent segments and edges relations between these segments. A relation is a triple (x, y, z) where x, y are similar segments and z represents their similarity. Based on the segment types, we distinguish three types of relations in this graph:

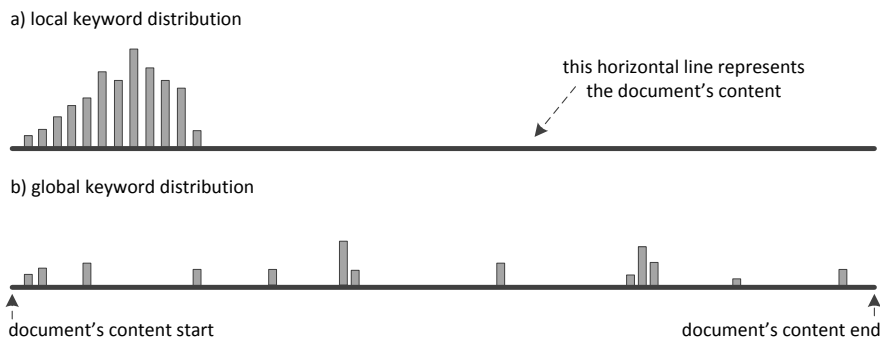


Figure 11.2. An example of two keyword distributions where a) is local and b) is global. It is clear that the distribution variance is higher in case of b). Therefore in this case we consider the keyword global

- *Parent relations*—exists between similar local and global segments coming from different documents (x is local segment, y global or vice versa).
- *Global relations*—exists between similar global segments (both x and y are global).
- *Local relations*—exists between similar local segments (both x and y are local).

Two segments are related if they share some minimum number of keywords. This value can differ for different types of relations. For now, the minimum similarity is 0.25 and was set based on the results of carried experiments.

Found relations are stored in a segment graph in a form of an edge connecting two segments. This edge is associated with a value ranging $[0, 1]$ which represents a relation strength. If two segments are related, they exchange their keywords. For instance, segment $S1$ contains *keywords goal, question, metric, GQM, success* and segment $S2$ *goal, question, metric, GQM, project*. These two segments are related, therefore, they exchange their keywords. The result after the exchange will be $S1$ *goal, question, metric, GQM, success, project* and $S2$ *goal, question, metric, GQM, project, success*. The whole situation is depicted in the Figure 3.

Experiments

This section describes the carried experiments with the proposed model. First we explain how we determined model parameters, e.g. text block length, tf-idf threshold. Then we show how well our approach performed when we compared its results to the output of plagiarism detection system PlaDeS described by Chudá, D. et al. (2011). PlaDeS is a plagiarism detection system developed at Slovak University of Technology, Faculty of Informatics and Information Technologies. We decided to use this system because it performs plagiarism check over documents written in Slovak language and because we know details about this system as we are its authors.

We had two datasets on which we carried out our experiments. The first dataset contains articles from web written in English language, the second one consists of student assignments written in Slovak language.

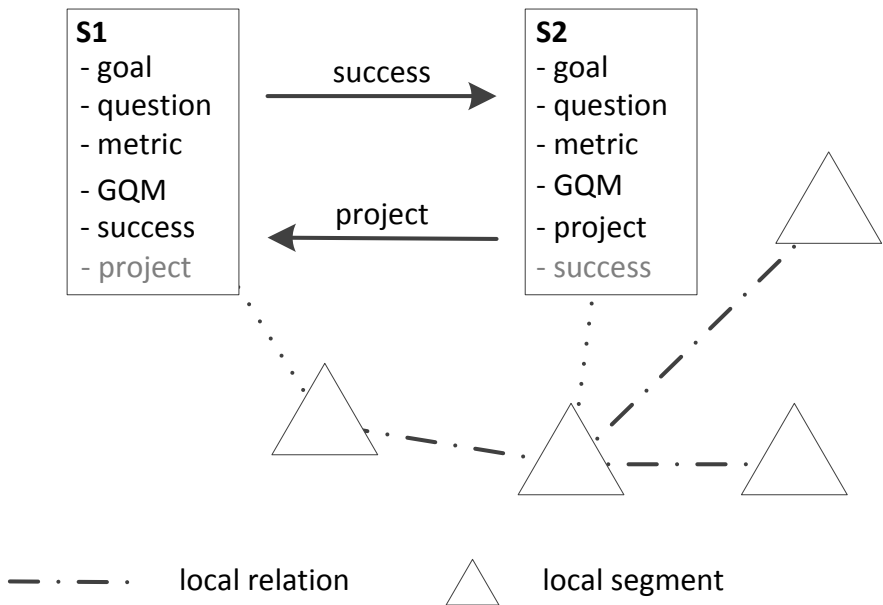


Figure 11.2. An example of a segment keyword exchange in the created segment graph. The keywords with lighter colour in the tables are those that were exchanged

Model parameter estimation

For the parameter estimation of the proposed model we used web articles from the BBC Travel. We selected 220 web pages, each web page dealing with exactly one topic, and manually annotated them. The annotated articles were about 600 words long on average. The output of the annotation process was a set of at least 5 to max 10 keywords associated with every web article.

Here is the list of the parameters that we estimated empirically on this dataset. The number after the hyphen represents the estimated parameter value for which we achieved the best results.

- text block (TB) length in letters—260
- tf-idf threshold—0.03
- minimum number of keywords in a local segment—3
- maximum number of keyword in global segment—6
- minimum threshold for local-to-local segment similarity in order to be related—0.3
- minimum threshold for global-to-global segment similarity in order to be related—0.3

To estimate these parameter values we used our approach to find all global keywords for every web article and compared the output with the user's keywords (user's

keywords are those, that were the output of the annotation process). As a global keyword we considered top 5 keywords with the highest variance returned by our method. The tf-idf approach achieved precision 0.40 and recall 0.27. Our approach achieved precision 0.46 and recall 0.30. Clearly our approach outperformed the tf-idf in the keyword extraction process.

Performance

To evaluate the overall performance of our approach, we compared its output to the output of a plagiarism detection system PlaDeS. As a detection method in this system we used 3-grams.

PlaDeS returns similar pairs of documents as triples $(d1, d2, s)$, where $d1$ and $d2$ are the compared documents and s represents their similarity. The returned similarity can be different considering the order in which the documents are compared. For instance, the s value for two triples $(d1, d2, s)$ and $(d2, d1, s)$ will differ in most cases (asymmetric similarity). However, our approach that we proposed in this paper does not distinguish the order of the compared documents. Remember that the relation is defined as a triple (x, y, z) , where both (x, y, z) and (y, x, z) return the same value z (symmetric similarity). Therefore, we compare the output of our solution with PlaDeS in such way, that if our approach identifies relation (x, y, z) , we check if the output of PlaDeS does not contain a pair (x, y, s) or a pair (y, x, s) . We do not search for a match in all document pairs returned by PlaDeS, but only those with similarity higher than 4%. This threshold was estimated based on our own experience with plagiarism checking.

Our dataset contained 313 student assignments written in Slovak language. An average length of each assignment was about 2500 words. The overall performance of our method is depicted in Figure 4. This figure shows the portion of detected suspicious document pairs (y axis) for different similarity thresholds (x axis) when compared to the output of PlaDeS. For instance, if we consider all document pairs returned by PlaDeS with similarity 20% (x axis), our approach detected 60% (y axis) of these pairs. The total number of unordered pairs of documents returned by PlaDeS was 48 828 while our approach detected 14,530 pairs. Figure 4 represents the achieved recall of our system when compared to PlaDeS. However, we were not able to evaluate the precision of our system with this experiment. This is because we cannot tell which document pairs returned by our method belong to the similarity pairs returned by PlaDeS (there is a different similarity calculation used in these two systems).

This high number (14,530) of document pairs detected by our approach is due to the large amount of identical topics. The used dataset consists of student assignments from last 8 years, while in each year there were around 15 different topics. The total number of students during these 8 years was 313. This means, that there can be found many possible similar pairs between students' assignments within different years. This could be one explanation why the carried experiments showed such results (Figure 4).

Conclusion

In this paper we presented a novel approach to plagiarism detection between documents written in natural language. This approach segments documents on subtopics by

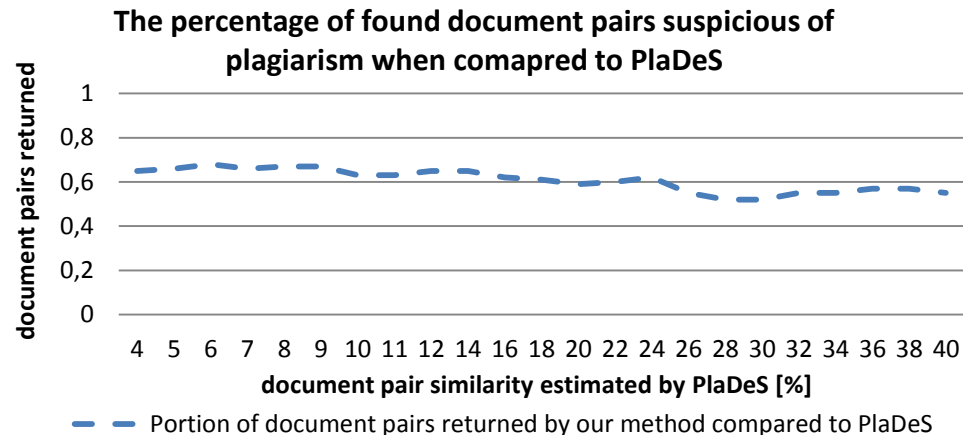


Figure 11.2. Recall of our method. Shows portion of detected suspicious document pairs by our method when compared to the output of a plagiarism detection system PlaDeS

watching keyword distributions in document’s content. From the identified segments we construct a segment graph that stores found relationships between these segments belonging to different documents. Based on these relations, document pairs with similar topics are extracted and recommended on special plagiarism checking. By special checking we mean computationally more complex comparison techniques than standard approaches.

The results showed that our approach detected about 60% of all suspicious document pairs returned by a plagiarism detection system PlaDeS. The total number of document pairs that our solution returned is 14 530 which represents the 29.76% of all possible pairs that have to be compared by PlaDeS. If we now wanted to use a more sophisticated method to compare some of the similar segments found in the documents, we would need less computation power than in the case of using PlaDeS

As a main advantage of our approach we consider the characteristics of the proposed algorithm that we use to find relations between segments belonging to different documents. Because we distinguish two types of segments, global and local, the relations between these segments also differ. Probably the biggest benefit this has is that it enables us to deeply explore the found similarities between documents. For instance, we can differently focus on analysing the local-to-global and global-to-global relations in the builded segment graph. Possible drawbacks of the proposed solution are that it is mainly suited for longer text documents. Although its performance on shorter texts has not been determined yet, we expect its drop. Also we have not yet determined what computation cost our solution has in case when comparing thousands of text documents.

In order to improve the plagiarism detection, in the nearest future we decided to integrate the described approach into the plagiarism detection system PlaDeS. We expect from this step to be able to better detect paraphrasing, because more complex approaches for plagiarism detection can be applied on related document segments.

For now, we did not take any advantage in the carried experiments from the different types of relations stored in the graph (local-to-local, global-to-global, etc.). In the future we would also like to use different analyses techniques with different types of relations. For instance, based on the type of found relations in segment graph we might be able to determine if students copy more local areas of documents or tend to copy documents as a whole.

References

- Ambwani, G., Davis, AR. (2010) Contextually-mediated semantic similarity graphs for topic segmentation. 2010. *Workshop on Graph-based Methods for Natural Language Processing*, 60–68 2010, Stroudsburg USA.
- Blei, MD., Ng, YA., Jordan, IM. (2003) Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Cedeño, BA., Rosso, P., Benedí, MJ. (2009) Reducing the Plagiarism Detection Search Space on the Basis of the Kullback-Leibler Distance. *10th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing '09)*, 523–534 2009, Berlin Germany.
- Chudá, D., Kučečka, T. (2011) PlaDes—Effective Way to Plagiarism Detection of Student Assignments Written in Natural Language. *1st International Conference on E-learning and the Knowledge Society (E-learning'11)*, 47–52 2011, Bucharest Romania.
- Hofmann, T. (1999) Probabilistic latent semantic indexing. *22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 50–57 1999, New York, USA.
- Kučečka, T. (2012) Identifikácia previazania slov prostredníctvom ich rozloženia v dokumente [In English: Identifying Keyword Relations Based on Their Position in Text]. *7th Workshop on Intelligent and Knowledge Oriented Technologies*, 45–48 2012, Smolenice Slovakia.
- Kučečka, T., Chudá, D. (2012) Identifikácia témy a podtémy dokumentu na základe pozície kľúčových slov v texte [In English: Topic Identification Based on Keyword Positions]. *Annual Database Conference (DATAKON)*, 127–136 2012, Czech Republic.
- Sun, B., Mitra, P., Giles, CL., Yen, J., Zha, H. (2007) Topic segmentation with shared topic detection and alignment of multiple documents. *30th annual international ACM SIGIR conference on Research and development in information retrieval*, 199–206 2007, New York, USA.

Acknowledgements

This work was partially supported by the Scientific Grant Agency of the Slovak Republic, grant No. VG1/0971/11 and is the partial result of the Slovak Research and Development Agency under the contract No. APVV-0208-10.

Authors

Tomáš Kučečka, kucecka@fiit.stuba.sk,
Daniela Chudá, chuda@fiit.stuba.sk
Slovak University of Technology, Slovakia

Copyright © 2013 Authors listed on page 261: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

PREDICTING NUMBER OF SEARCH ENGINE RESULTS TO OPTIMISE ONLINE PLAGIARISM DETECTION

Ondřej Veselý, Jan Kolomazník

Abstract: At our university has developed a new system for online plagiarism detection. The development version uses the search engine which is exposed to loads of search queries with suboptimal length. The aim of the research is to minimise the number of queries and keep the detection sensitivity on the same level. The paper presents a few methods of determining the number of search engine results for given query. The aim is to estimate the optimal phrase to be send to the search engine. Estimations were confirmed by neural network; after that a statistical model was created and tested. The model is fitted with data extracted from massively processed language corpus, where normalised Google distance for each common bigram is used for creating a query-specific context vector which characteristics was found to be related to the number of result for given query.

Introduction

Today's systems for plagiarism detection (hereinafter called as antiplagiaters), which use search engines (SE) for online plagiarism detection, deals with problem of determining optimal search query size. The usual technique is based on the process of selecting some phrases from the given document and adding the list of found web address for each query to the result set for next analysis. Potthas (2012) calls different phases of this process as chunking, keyphrase extraction and query formulation. The main problem is how to determine the number of words which should be included to the search query. Usually a static length of a fragment of text is used to create a query (for example in Copyscape checker service). According Nina, Peter and Wilkinson (2007) it has the right query length is strongly correlated to the "broadness of generality" in the context of query's meaning. We believe that using a variable length fitted to the meaning (called *optimal query length*) improves the sensitivity of detection which is discussed here but it's confirmation is not covered by this paper. The aim of this paper is to present it's optimisation.

When using longer than optimal phrases, only a minor modification of tested plagiarised document can make it undetectable. The advantage is that the detection itself is faster, because the ratio between whole number of query and document size is low.

Use of shorter than optimal phrases has two disadvantages—the number of queries is relatively higher and there is strong dependency on search engine algorithms for sorting the results. Low number of words in a search query usually leads to huge numbers of search results, which are inaccessible without more listing queries.

The most precise results can be obtained when the phrase size is determined dynamically by querying the SE with increasing length of selected phrase and use the last non-empty result set (Veselý, 2013) Veselý and Kolomazník (2012) call that a *naive approach*.

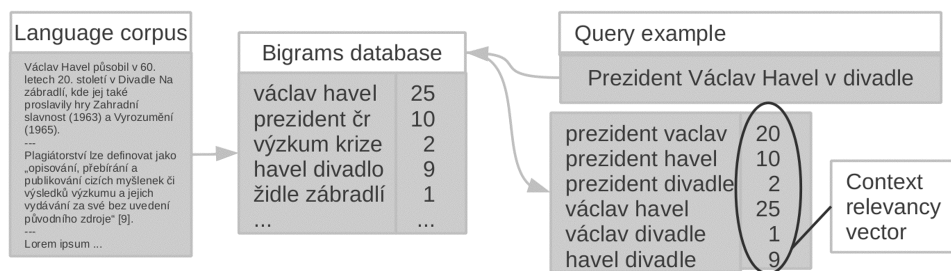


Figure 11.2. Creation of datasets of bigrams where the context similarity for each pair of words is stored to retrieve context relevancy vector for an arbitrary query

The disadvantage is the huge number of queries to the SE; one query for every tested word.

There was an unsuccessful attempt for optimisation based on presumption that “there is a relation between vector of frequency occurrences of each word in query and the number of results we got from the search” (Veselý, Kolomazník 2012), but the conclusion was clear “neural network has not reveal any relation between the frequency of occurrence of the word in the query and the number of results of search engine”. Due to the fact that a similar principle is widely used on Google Similarity Distance (Cilibrasi, Vitanyi, 2007) applications, we presume that for successful estimation of SE results is necessary to account the mutual context of the words, not only its frequency occurrence. It leads to the necessity of massive pre-processing of language corpus to get a similarity index for each common pair (bigram) of words. This article is to verify the hypothesis that the set of these indexes is the key for determining the number of search results to optimise plagiarism detection.

Methods

There is a brief list of steps we followed to conduct the experiment:

1. creation of two large datasets of bigrams where the context similarity for each pair of words is calculated and stored,
2. implementation an bigrams database for determination of context relevancy vector for an arbitrary query,
3. use of neural network and statistical regression to interpret the context relevancy vector so the estimation of number the SE results can be calculated for related query,
4. testing the accuracy of estimation with the SE data.

The first two steps are represented by following schematic.

Two different datasets were created because two difference functions were needed. One for estimating the number of SE results by quoted queries (which means that the query represents the exact phrase we are looking for; it's also called phrase search) and another dataset for unquoted queries.

Table 1

Two sets of interval used for context vector normalisation

Class	1	2	3	4	5
No quotes intervals	0–1	1–2	2–4	4–11	11+
Quotes intervals	0–1	1–3	3–11	11–1000	1000+

The Czech National Corpus (SYN2010) was used for estimating the number of results for phrase searches. All pairs of adjacent words from the corpus were added to the dataset; if the pair was already in the set, the related counter was increased. This resulted in a gigabyte file where a pairs of words were associated with the value of the occurrence counter (see table bigrams database on fig. 1)

We did not have direct access to the corpus which we needed in order to create the next dataset for the unquoted corpus, so an alternative corpus (Majliš, 2012) was used. In this case we indexed all words in each corpus fragment, not only the adjacent ones, in order to reflect the SE behaviour. This led to the exponential increase of time and space complexity; the procession itself takes four weeks on a common computer in spite of the fact that we ignored stop words and words shorter than 5 characters. Also a really simple lemmatization was used: the last two characters of the words were removed to suspend the morphology difference for the words with the same meaning.

As we considered this as a prototype, we analysed only about one third of the corpus. Implementing comprehensive solutions like MapReduce would definitely improve the actually reached prediction level.

As described in steps 1 and 2 we constructed a system which could provide a context relevancy vector, which is believed to be related with the number of results, we need to find and formulate the relation. For confirmation of existence of the relation, the two layer neural network with backpropagation learning was used.

Step 3. is fulfilled by experiments described in the next sections. Two query datasets were used: quoted (Qq) and unquoted (Qw) queries together with the number of hits given by SE.

Creating and normalising the context vector

As it is showed in fig 1, the context vector is a simple set of context values from bigrams database for (almost) every pair of words in query. The short words are ignored. This fact results to variable length of the context vector exponentially related to the query length.

The process of vector normalisation was based on classifying each number in given vector to the intervals in table 1. The resulting normalised vector was constructed from number of occurrences of each class.

Table 1 shows the chosen distribution of intervals mapped to the position in the context vector. For example, normalised context vector from fig 1 is (1, 1, 0, 2, 2).

Table 2
Mean square error (MSE) value determined by neural network on different datasets

	by number of words in query	by context vector
Q_q dataset	26.0	18.5
Q_w dataset	45.0	41.0

Experiment 1, confirmation of the relation

Let Q_w be the unquoted query and Q_q the quoted query. Let $SE(Q)$ be a number of result of the SE for a given Q and $CV(Q)$ the context vector of Q .

We will try to fit a model $NN(CV(Q))$ which could confirm the following relation:

$$NN(CV(Q)) \sim \log_{10}(SE(Q))$$

In the other words, we want to provide functionality to estimate the magnitude of SE results for a given Q .

The following table compares resulting mean square errors when predicting $\log_{10}(SE(Q))$ from two query datasets (made by Q_w and Q_q types) both populated by two different calculations which were used as training data. The first calculation was based on the number of words in the query (as a method described in SYN2010). The second calculations consist of described context vector given a token from bigrams database. The context vector was normalised as described before.

Table 2 shows that with a context vector there is a noticeably lower MSE in both datasets. This means that calculation of the context vector contains the information which could be used to improve the number of SE hits in comparison to the simple estimation with a query word count.

The learned neural network is good enough for operational use, but the disadvantage of a neural network is it's black box like behaviour—we could not understand the relation deeply enough to satisfy the academic purpose. Only two layers of the neural network gave us a rational reason to speculate that the relation is simple enough to be described by a statistical model, so we made another experiment.

Experiment 2, finding the relation by creating statistical model

We expected, that relationship can be formally expressed by mathematical formula in order to have a tool to calculate $\log_{10}(SE(Q))$ from $CV(Q)$. To reveal the relation formula (one coefficient for each member of $CV(Q)$), we replaced neural network NN , by statistical linear model LM .

$$LM(CV(Q)) \sim \log_{10}(SE(Q))$$

The data from Experiment 1 was used for fitting LM implemented in R-project like *lm* function. The model determined the relation between the common logarithm of number of hits given by SE (for Q_w) and the elements of the context vector. For expected better performance the model was fitted for both type vectors—quoted (q_1, q_2, q_3, q_4, q_5) and unquoted (w_1, w_2, w_3, w_4, w_5).

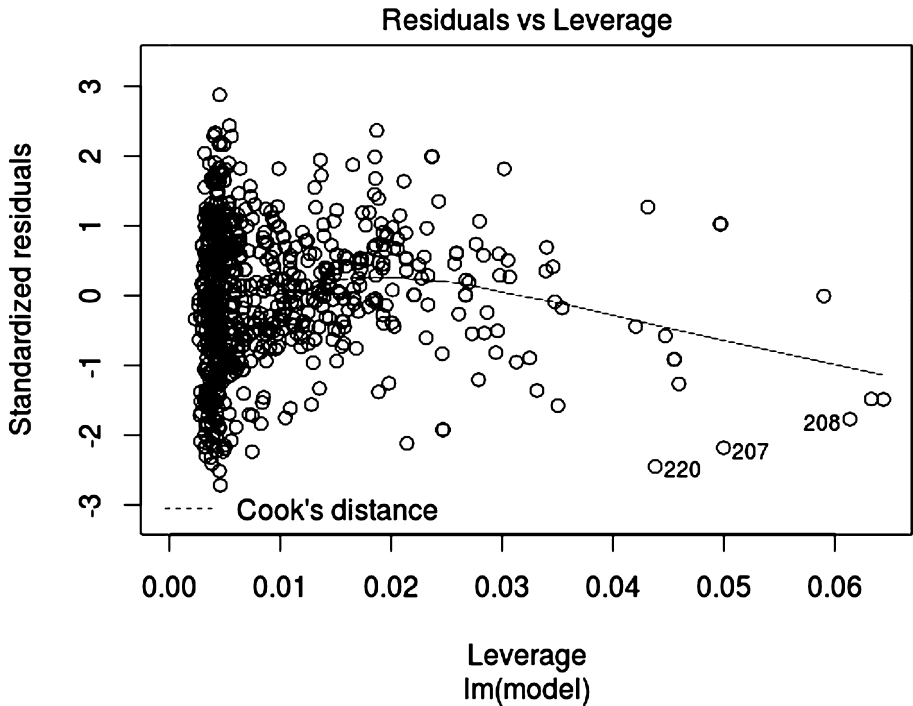


Figure 11.2. Distribution of standardised residuals in the fitted model

Table 3

The accuracy of the statistical model when for estimation the magnitude of SE results. Each column represents the estimation difference $\Delta e = \text{abs}(\text{LM}(\text{CV}(Q)) - \log_{10}(\text{SE}(Q)))$

	$\Delta e = 0$	$\Delta e = 1$	$\Delta e = 2$	$\Delta e = 3$
% of queries	39.514978602	30.5278174037	18.1169757489	11.8402282454

The fitting found these elements relevant: $(q_1, q_4, w_1, w_2, w_4, w_5)$; the coefficients found are showed in the following polynomial function.

$$\text{LM: } y = -0.4847 \cdot q_1^2 - 0.1091 \cdot q_4 - 0.3983 \cdot q_1 + 0.0795 \cdot w_1 + 0.2755 \cdot w_4 + 0.0809 \cdot w_5$$

Obviously, the q_1 element has the strongest negative influence to the result. It means that there are many words in the query with no mutual context. The positive coefficients for almost all elements in the unquoted vector probably signal the insufficient size of the bigrams database.

Table 4
Example: Estimation the SE results for different length of queries and determinig the new order of querying the SE

words	Query Q	LM(Q)	New query queue order
8	prezident václav havel v divadle na provázku uvedl	3	5.
9	prezident václav havel v divadle na provázku uvedl famózní	2	3.
10	prezident václav havel v divadle na provázku uvedl famózní vystoupení	1	1.
11	prezident václav havel v divadle na provázku uvedl famózní vystoupení bratrů	1	2.
12	prezident václav havel v divadle na provázku uvedl famózní vystoupení bratrů Nejezchlebových	0	4.

The provided function gives us the possibility to estimate the common logarithm of the number of SE hits in for almost every second query.

Experiment 3, benefits for plagiarism detection

We built a statistical model LM to enhance the naive algorithm for determining the optimal length of query. The function lm in the statistical modelling tool “R” was used. Instead of gradually increasing the number of words in query, querying the SE in every run of the loop, we used the model which arranges the given query length possibilities by the probability of being estimated as the optimal ones. In the other words, the new heuristics algorithm prefers the queries Q where $LM(CV(Q)) \rightarrow 1$ so, presumably the $\log_{10}(SE(Q)) \rightarrow 1$ which so $SE(Q) \rightarrow 10$ (as we supposed this value to be considered as “optimal enough”).

We ran antiplagiator system desribed by Veselý (2012) twice on the 76 different queries extracted from a common diploma thesis text. First run was made with naive algorithm, the second one used the heuristics described in this section.

Results

The usage of the neural network confirms the strong relation between the context vector generated from bigrams database and the number of SE results for given query; both quoted and unquoted.

The statistical model LM described in previous section provides us a heuristics allows the antiplagiator to find the optimal length of the query with less number of suboptimal queries.

On the dataset consists of 76 unquoted queries, we ran a benchmark which compares the chosen metrics—ratio between the number of suboptimal queries and the found optimal ones.

Table 5
Table shows the number of suboptimal requests need to found the desired number of optimal queries. Two algorithms are presented

	The number of all request	Optimal queries found	ratio
Naive algorithm	683	76	8.986842
LM heuristics used	345	76	4.539474

Described estimation of the SE results allow us use the SE only with a less than a half of a number of queries needed to determine the optimal queries during the detection.

Discussion

Despite the benefit of the speed-up provided by the resulting functionality, there is still questions about how important is to work with optimal queries and how strong could be the influence of usage of suboptimal queries instead. It would definitely degrade the sensitivity of the detection, but the another speed-up could be a better compromise.

The term speed-up is used in the context of decreasing the number of requests to SE. But the process requires another kind of requests to bigrams database to get the context vector. For now, the composition of context vector takes about a minute which is not definitely a “speed-up” in the context of time. However it is mainly a technical problem which could easily solved by moving the whole database into random-access memory.

Another question that we did not answer concerns the efficiency of the system in case we would process the whole corpus to make a comprehensive bigrams database. Although the chosen distribution of the intervals in table 1 could be replaced by more sophisticated set which would better reflect the distribution of the values.

This questions are expected to be answered at 10th international competition on plagiarism detection “PAN 2014”, when the English corpus will be ready and system market as full capable to participate in evaluation labs.

Determination of a context vector may be used in a broader context than plagiarism detection. As a “commonness function” for any piece of text, it can serve as detector of computer generated dummy text without meaning, or possibly as a semantic detector of chapter borders.

Conclusion

The article described a process of creating semantic characteristics called context vector, which determines the context of similarity for any set of words in specific language. The experiment shows an application of the context vector for estimating the number of SE results for given phrase. The efficiency of finding the optimal length of query in task of plagiarism detection increased two times compared to the previous naive approach.

References

- Cilibrasi, RL., Vitanyi, PMB. (2007). The Google Similarity Distance, IEEE Trans. Page 370–383. Knowledge and Data Engineering
- Majliš, M., Žabokrtský, Z. (2012) Language richness of the web. In *Proceedings of LREC2012*, Istanbul, Turkey, May 2012. ELRA, European Language Resources Association
- Phan, N., Bailey, P. & Wilkinson, R. (2007) Understanding the Relationship of Information Need Specificity to Search Query Length. *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 709–710. New York, NY, USA, ACM
- Potthast, M., Gollub, T., Grassegger, J., Kiesel, J., Michel, M., Oberländer, M. & Tippman, M. Overview of the 4th International Competition on Plagiarism Detection. In: *CLEF 2012 Evaluation Labs and Workshop* [online]. Rome, Italy, 2012 [cit. 2013-04-30]. ISBN 978-88-904810-3-1 ISSN 2038-4963. Accessed from: http://www.uni-weimar.de/medien/webis/publications/papers/stein_2012t.pdf
- SYN2010. Ústav Českého národního korpusu FF UK, Praha 2010. Available from <http://www.korpus.cz> [accessed 24 February 2013]
- Veselý, O., Kolomazník, J., Foltýnek, T. (2012) Heuristic and AI approach to optimize plagiarism detection tool using a public search engine. [CD-ROM]. In *IADIS International conference WWW/Internet*, pp. 309–403. ISBN 978-989-8533-09-8
- Veselý, O. (2013). *Analýza podobnosti závěrečných prací a on-line dokumentů*. Brno, CZ. Mendel University in Brno

Acknowledgements

The research was funded by the project Impact of Policies for Plagiarism in Higher Education Across Europe (510321-LLP-1-2010-1-UK-ERASMUS-EMHE).

Authors

Ing. Ondřej Veselý (ondrej.vesely@orwen.org),

Ing. Jan Kolomazník (jan.kolomaznik@mendelu.cz),
Mendel University in Brno, Czech republic

Copyright © 2013 Authors listed on page 270: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

WORKSHOPS, POSTERS

YES, THEY ARE ALL PLAGIARISING, BUT WHAT CAN BE DONE TO STOP IT?

Irene Glendinning

Abstract:

This paper supports a workshop about strategies for discouraging student plagiarism and academic dishonesty. The session will draw on examples of good practice revealed in the course of recent research and encourage participants to contribute their own experiences and ideas.

The author has witnessed diverse and interesting reactions to the revelation that she is involved in a project about plagiarism. Most researchers in education believe this is indeed a serious problem and something needs to be done. As the title of this paper implies, several respondents have expressed despair about what could or should be done about it. However a few otherwise well informed European educationalists have been genuinely at a loss to know why the project is needed and asked what is meant by the term plagiarism. Such anecdotes help to explain why the project *Impact of Policies for Plagiarism in Higher Education Across Europe* (IPPHEAE) was considered important enough to be funded by the European Union.

The three-year project is now in the final year of operation and the first results are due for dissemination in June 2013. The research includes a very wide ranging survey across the 27 EU countries, but also aims to document some interventions and case studies. The focus of the research is to explore policies in HEIs across EU countries for addressing student plagiarism at bachelor and master's level, to determine how consistently a learning community recognises the institutional policies and to capture some measure of how effective they are.

The workshop supports a project objective to share examples of good practice and success stories. But additionally the results from this session will contribute to the body of knowledge already amassed from the IPPHEAE project research, bringing together additional ideas and experiences from interested participants, many of whom will be actively researching into aspects of academic integrity.

Key words: Plagiarism, academic Integrity, Higher Education, prevention of plagiarism, IPPHEAE project

Workshop format—Discouraging plagiarism

The workshop will be very largely interactive and participative.

Main topics:

- Presentation of some IPPHEAE project findings: Worst cases—best cases, known unknowns and unknown unknowns;
- Exploring and unpacking some examples of good practice;
- Deciding where to begin in countries and higher education institutions where:
 - No policies and procedures for plagiarism and academic dishonesty are being implemented at present;
 - There is no appreciation of the scale and nature of student plagiarism;
 - There is a culture of academic autonomy;
 - Staff development is unheard of;

- Whistle-blowers on plagiarism are seen as undesirables;
 - Plagiarism prevention or avoidance is not understood.
- Summarising recommendations for developing an EU-wide approach to excellence in academic integrity.

Author

Irene Glendinning, Academic Manager for Student Experience, Faculty of Engineering and Computing, Coventry University, UK

Irene Glendinning is Academic Manager for Student Experience for the Faculty of Engineering and Computing at Coventry University. She is Principal Investigator for the European Union Lifelong Learning Project “Impact of Policies for Plagiarism in Higher Education Across Europe” 2010–2013, funded by the European Commission. Her research interests include quality assurance, student diversity and equality, pedagogy, student employment and employability. She has been an academic in Higher Education since 1990, with extensive experience of programme management covering a range of disciplines and lecturing in computer science. Earlier in her career she was a systems analyst and computer programmer and she also taught mathematics, IT and computing at secondary school level.

ANALYSIS OF THE PERCEPTIONS OF UNDERGRADUATE STUDENTS IN BUSINESS ADMINISTRATION ON THE OCCURRENCE OF ACADEMIC PLAGIARISM IN BRAZIL

Marcelo Krokosz, Reinaldo Putvinskis

Abstract: Several studies in academic literature analyze undergraduate students' perceptions about plagiarism (Power 2009), related factors (Curtis & Popal 2011) and measures to avoid it (McCabe et al. 2001), among other issues. However, a review of Brazilian literature on plagiarism indicates that knowledge on this subject is still very rudimentary (Krokosz 2012), and it has been observed that efforts to combat plagiarism are practically nonexistent at Brazilian universities (Krokosz 2011).

The purpose of this study is to expand knowledge about plagiarism in Brazil; perceptions regarding plagiarism were identified and analyzed using a sample of 373 undergraduate students from five universities. The participants answered a questionnaire containing 25 questions about their understanding (definition, types), forms (accidental or intentional) and reasons for plagiarism and how to avoid it.

The results obtained are similar to others found around the world. The reasons given for plagiarism were: lack of time, the desire to obtain good grades and difficulties with academic writing. Although a majority has demonstrated a clear understanding of verbatim plagiarism, most disagreed that pastiche is plagiarism and some respondents demonstrated lack of understanding about how to paraphrase. Furthermore, there is evidence that not all types of plagiarism are recognized by students. We discuss the need for the discernment of international strategies for avoiding plagiarism in order to adopt those that are broadly applicable in Brazil, and recommend the development of other studies involving professors and institutions in order to develop appropriate measures for dealing with plagiarism in the Brazilian context.

References

- Curtis, G.J. & Popal, R. (2011) An examination of factors related to plagiarism and a five-year follow-up of plagiarism at an Australian university. *International Journal for Educational Integrity*, 7(1), 30–42
- Krokosz, M. (2012) A literature review of scientific research and reflections on plagiarism in Brazil since 1990. In *5th International Plagiarism Conference 16–18 July 2012*, Newcastle UK. Available from http://archive.plagiarismadvice.org/documents/conference2012/posters/Krokosz_poster2.pdf [accessed 4 April 2013]
- Krokosz, M. (2011) Abordagem do plágio nas três melhores universidades de cada um dos cinco continentes e do Brasil. *Revista Brasileira de Educação*, 16 (48), 745–818. Available from <http://www.scielo.br/pdf/rbedu/v16n48/v16n48a11.pdf> [accessed 4 April 2013]
- McCabe, DL., Treviño, LK. & Butterfield, KD. (2001) Cheating in Academic Institutions: A Decade of Research. *Ethics & Behavior*, 11(3), 219–232
- Power, LG. (2009) University Students' Perceptions of Plagiarism. *Learning*, 80 (6), 643–662

Authors

Marcelo Krokosz, marcelix@usp.br

Reinaldo Putvinskis, profreinaldo@fecap.br

University of São Paulo and FECAP, Brazil

Copyright © 2013 Krokosz, Marcelo and Putvinskis, Reinaldo: The authors grant to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the authors.

ACADEMIC DISHONESTY AND EXTENUATING CIRCUMSTANCES IN THE GLOBAL CLASSROOM

Trudy Somers

Abstract: Detected plagiarism in the global classroom may be treated as ethical breach (dishonesty or theft), lack of skill (in language or writing), ignorance (of cultural definitions or expectations of source usage and acknowledgement), or error in judgment. Many institutions offer or require workshops for newly matriculated students about the standards and consequences of academic integrity violation. These workshops, preventive measures, range in content from provision of penalty lists to exploration of culturally nuanced academic integrity issues. Institutions may provide opportunity for remediation after a detected plagiarism incident. The course instructor, as first responder, often provides detection, prevention and remediation. Quality, consistency, and institutional reputation are at stake. This workshop affords participants the opportunity to discuss various examples of plagiarized material and appropriate (or inappropriate) actions. Workshop participants will be divided into small groups for discussion of a number of cases: an admissions application personal essay; a comprehensive doctoral exam; part of a PhD dissertation; and, essay assignments for undergraduate and graduate level courses. Discussion topics will include severity of the violation, extenuating circumstances and implications for prevention, detection, remediation, and penalty.

Workshop: Academic dishonesty and extenuating circumstances in the global classroom

This workshop affords participants the opportunity to discuss various examples of plagiarized material and appropriate (or inappropriate) actions. Workshop participants will be divided into small groups for discussion of a number of cases: such as, an admissions application personal essay; a comprehensive doctoral exam; part of a PhD dissertation; and, essay assignments for undergraduate and graduate level courses. Discussion topics will include severity of the violation, extenuating circumstances and implications for prevention, detection, remediation, and penalty. The workshop will begin with a brief discussion of a checklist for factors to consider when evaluating a potential academic integrity violation. (See Table 1 and the following discussion.) Then the cases provided in this paper will be discussed. At that time, participants will collectively choose to

- Continue discussion of additional cases (handouts will be provided)
- Explore additional factors that influence academic integrity in the global classroom and the challenges facing international students
- Examine various techniques used to provide training to students and faculty about plagiarism (handouts will be provided)
- Discuss various university plagiarism policies.

The session will conclude with a review of the topics covered an assessment of the levels of consensus reached. The most valuable take-away from this session will

Table 1

Factors to consider in plagiarism evaluation

Factor	Significant	Rank
Writing skill		
Language proficiency		
Cultural differences		
Plagiarism rules awareness		
Program level		
School policy		
Occurrence		
Other		

be hearing the variety of opinions from the other participants. Agreement on all points would not be reached, but the factors considered will be evaluated for range in consideration and weight. Action items may be formulated for participants to take back to home institutions or consortiums.

All plagiarism is not equal. Although universities must be fair, differences in students and other extenuating circumstances dictate that fairness is not identical treatment for all cases. Detected plagiarism in the global classroom may be an ethical breach, as students deliberately take work from another and claim it as their own (Baker, Berry and Thornton 2011). However, some students explain that it is a sign of respect to use, to imitate, the work of another (Pennington 2010). Other students resort to copying work out of frustration with their proficiency in the language of instruction, or their skill in writing that language (Ford and Hughes 2011, Pfannestiel 2010). A number of students, native speakers as well, profess ignorance of cultural definitions or expectations of source usage and acknowledgement (Kwong et al. 2010, Ford and Hughes 2011, Pennington 2010). Students also make errors in judgment by not allowing enough time to do an assignment or copying because an assignment was too hard, or too boring, or not realizing they might get caught (Olafson et al. 2012, Stevens et al. 2010).

Many institutions offer or require workshops for newly matriculated students about the standards and consequences of academic integrity violation to reduce plagiarism (Craig et al. 2010). These workshops and related preventive measures, vary in intent and content. Some explore culturally nuanced academic integrity issues in a broad ethical context (Alcota et al. 2012). Some explain the nature of text matching services (Hariharan) while others rely on electronic monitoring and surveillance (Kitahura) Institutions may provide opportunity for remediation after a detected plagiarism incident. Materials to discuss these formats will be available in the workshop. There will be a handout identifying various offerings.

References

- Alcota, M., de Gauna, PR., & Gonz lez, FE. (2012) Development of ethical practices and social responsibility in Dental Education at the University of Chile: Student and Faculty Perceptions. *European Journal of Dental Education*
- Baker, RK., Berry, P., & Thornton, B. (2011) Student attitudes on academic integrity violations. *Journal of College Teaching & Learning (TLC)* 5(1), 5–14
- Craig, PA., Federici, E., & Buehler, MA. (2010) Instructing students in academic integrity. *Journal of College Science Teaching* 40(2), 50–55
- Ford, PJ., & Hughes, C. (2011) Academic integrity and plagiarism: Perceptions and experience of staff and students in a school of dentistry: A situational analysis of staff and student perspectives. *European Journal of Dental Education* 16(1), e180–e186
- Hariharan, S., Kamal, S., Faisal, AV., Azharudheen, SM., & Raman, B. (2010) Detecting plagiarism in text documents. *Information Processing and Management* 497–500.
- Kitahara, R., Westfall, F., & Mankelwicz, J. (2011) New, multi-faceted hybrid approaches to ensuring academic integrity. *Journal of Academic and Business Ethics* 3, 1–12
- Kwong, T., Ng, H., Mark, K. & Wong, E. (2010) Students' and faculty's perception of academic integrity in Hong Kong. *Campus-Wide Information Systems* 27(5), 341–355
- Olafson, L., Schraw, G., Nadelson, L., Nadelson, S., & Kehrwald, N. (2012) Exploring the judgment-action gap: College students and academic dishonesty. *Ethics & Behavior*, forthcoming
- Pennington, MC. (2010) Plagiarism in the academy: Towards a proactive pedagogy. *Writing & Pedagogy* 2(2), 147–162
- Pfannenstiel, AN. (2010) Digital literacies and academic integrity. *International Journal for Educational Integrity* 6(2), 41–49
- Stephens, J., Romakin, V., and Yukhymenko, M. (2010) Academic motivation and misconduct in two cultures: A comparative analysis of US and Ukrainian undergraduates. *International Journal for Educational Integrity* 6(1), 47–60

Cases for workshop discussion

Case 1

The professor began reading a paper submitted by Zoe, her fourth for the course. Although the course was graduate level, Zoe had struggled throughout with writing in English (not her native language) and following the required APA format. The professor referred Zoe to the university writing center after her first submission. They had painstakingly worked through grammar, punctuation, word choices and other mechanics. The professor gritted her teeth and began to read aloud (that sometimes helped figure out word meanings) and was delighted to find an engaging, well written manuscript. However, the Turnitin report (a requirement for that assignment) showed that the document was copied from another paper, nearly verbatim. What factors should be considered? What action should be taken?

Case 2

Simcha was writing her dissertation proposal in the area of women and leadership for the PhD degree. Her chairperson was annoyed because repeated attempts to correct concept issues, grammar and spelling had been ignored, for the most part. This new paper was different. It was conceptually sound and well written, with a new focus on emotional intelligence in female leaders. The professor was delighted until a phrase stuck in his mind. It was an unusual way to refer to emotional intelligence that a former student had developed about a year earlier in his dissertation proposal. The professor searched through old documents and found the paper. A comparison revealed startling similarities in the two documents. It looked like a simple search and replace operation had substituted “female” for the industry group mentioned in the earlier document. Simcha acknowledged the wisdom of this former student in another document and expressed her gratitude for his contribution to her work. What factors should be considered? What action should be taken?

Case 3

The professor submitted Maks’ third assignment of the fundamentals course—an introduction to university studies—to a text matching service. The results were returned to Maks: 30% of the material was attributable to several other sources. Some inaccurate citations were provided. A few paragraphs were copied wholesale without acknowledgement. Although paraphrasing was attempted, changes were not sufficient, even when the work was cited. The professor provided the report to the student with a detailed discussion of major violation areas. Part of the next assignment asks Maks to take the university academic integrity tutorial, to reflect on this third assignment, and to sign an honesty pledge. Should further action be taken at this time?

Cases 4–6

Will be provided at workshop.

Author

Trudy Somers, tsomers@ncu.edu, Northcentral University, USA

Copyright © 2013 Trudy Somers grants to the IPPHEAE 2013 Conference organisers and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to Mendel University in Brno, Czech Republic, to publish this document in full on the World Wide Web (prime sites and mirrors) on flash memory drive and in printed form within the IPPHEAE 2013 conference proceedings. Any other usage is prohibited without the express permission of the author.

ACADEMIC INTEGRITY IN HIGHER EDUCATION: A GLOBAL PERSPECTIVE

Tracey Bretag, Gill Rowell and Teddi Fishman

Abstract:

Speakers on this panel represent the International Association of Academic Integrity Conferences (IAAIC), an alliance of key academic integrity and plagiarism conferences worldwide. The IAAIC was formed to facilitate international conversations on educational issues ranging from cheating and plagiarism to pedagogy and best practices, and was launched in July 2012 at the 5th International Plagiarism Conference.

Panellists will begin by sharing some current best practices in academic integrity in Australia, North America and the United Kingdom. Then they will initiate a dialogue with conference participants about how to adapt and implement various education-focused approaches to promoting integrity, discouraging cheating, and addressing issues such as plagiarism and collusion in a variety of specific contexts.

Conference delegates will be invited to share their experiences with the panel on topics such as approaches to engaging students in meaningful discussions on academic integrity, innovative assessment strategies beyond the essay and how to establish robust and transparent institutional policies and procedures.

Key words: academic integrity, plagiarism, best practice

APPLYING PRINCIPLES OF SCHOLASTIC HONESTY AT VYSOKÁ ŠKOLA MANAŽMENTU V TRENČÍNE

Peter Kročtý

Abstract:

Vysoká Škola Manažmentu v Trenčíne/City University has more than 10 year experience in the area of scholastic honesty and application of zero tolerance to the acts of dishonesty by its students. During the last decade, our school in cooperation with its mother university City University of Seattle has been in a process of continuous monitoring and reviewing the process of applying policies and procedures to fight acts of scholastic dishonesty. All these steps are aiming at building a name of an institution whose graduates are people with high ethical standards. Even though we can claim based on results of the work of our scholastic honesty committee that there is a negative trend in the number of cases where scholastic dishonesty was found, there is still a space for improvement in practice and raising awareness of academic integrity among students and staff. This paper will provide an overview of the steps that have been taken with the recent assessment findings and suggest possible course of action for improvement in this area.

Key words: scholastic honesty, plagiarism detection, raising awareness of academic integrity

ON ANTI-PLAGIARISM IN AN OPEN UNIVERSITY: CHALLENGES AND BEST PRACTICES

G. Matheou, A. I. Kokkinaki and C. Demoliou

Abstract:

This case study aims to identify perceptions, practices and policies for avoiding plagiarism at a Tertiary Educational Institute in Cyprus that offers all programs at all levels using the methodology of distance learning. For the purposes of this case study, it will be referred as Libre University Cyprus (abbreviated to LUC). We examined how faculty, students and administrative staff at LUC perceive the term plagiarism, what cases of plagiarism have been identified, the tools and techniques followed for identification of plagiarism and academic dishonesty.

All modules at RUC are offered through an eLearning platform (eClass) to which students' assignments and term projects are being collected. Faculty members access the submitted work, provide comments and assign the grade. Students are required to have exams in person and their grade is a weighted average of the work submitted during the semester and the final exam.

In such an environment the issue of plagiarism and academic integrity overall undertakes new, interesting twists. This has been identified by University's bodies and the measures taken are above the national average. Based on interviews with students, we posit that several incidences of plagiarism and academic dishonesty have gone unidentified. If a tutor identifies or suspects a case of plagiarism, the tutor has to inform the module and academic coordinator. The student is invited to apologize for the plagiarism and the penalty depends on the seriousness of the plagiarism. If a student has been proved to plagiarize twice or more, then the case is referred to the Dean and then to the University's Governing Board. The student might get expelled, although no standards have been set and no precedence has been set. The small general paragraph on plagiarism in the University's Rules and Regulations does not provide enough guidance.

Tutors act individually depending on the seriousness of each case. There are cases of plagiarism that have not been reported by tutors. Reported cases may be handled by the Academic Council of the Program of Study and some more severe cases of academic dishonesty are referred to the Governing Board. There was a case that was so serious that the Governing Board decided to expel the student. For less serious cases the penalty the discretion of the tutor and can be reducing the degree or failing the assignment, at hand. In most programs of study, the tutor will decide if he/she will inform the Module Coordinator. On other programs of study tutor has to inform the Module Coordinator, even if the plagiarism is considered minor. There are programs of study that are consistently more flexible with the penalties than others, especially at the undergraduate level.

On conclusion, there are no procedures and standard penalties on plagiarism at LUC. Existing policies and procedures differ for almost every program of study. As a final note it is worth mentioning that there is no national legislation on the topic.

PLAGIARISM ACROSS EUROPE AND BEYOND
Conference Proceedings

First edition

Typeset by Jiří Rybička

Printed by MENDELU Publishing Centre

Brno 2013

ISBN 978-80-7375-765-6