Journal of English for Academic Purposes 11 (2012) 125–133

Contents lists available at SciVerse ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Gauging the effectiveness of anti-plagiarism software: An empirical study of second language graduate writers

Paul Stapleton*

Department of English, Hong Kong Institute of Education, 10 Lo Ping Road, Taipo, Hong Kong

Keywords:
Turnitin
Plagiarism
Anti-plagiarism
L2 writing
Academic writing

A B S T R A C T

The use of anti-plagiarism services has grown very quickly in recent years to the point where over half of American universities now have a license. The most popular of these services, Turnitin, claims that it is licensed in 126 countries and available in 10 languages suggesting that the service is becoming widely used around the world. In order to assess the effectiveness of this service, the present study compares the writing behavior of students in two equivalent classes, one of which was aware that their essays would be assessed for originality, while the other was not.

Results revealed that the class which was unaware of Turnitin had significantly higher rates of matching text, near copies and intentional plagiarism than the class which was aware of Turnitin. This finding suggests that Turnitin had a deterrent effect on plagiarism. However, while the raw percentages generated by Turnitin did provide a reasonable overall estimation of plagiarism, they did not always accurately reflect the extent of intentional plagiarism. The study concludes that anti-plagiarism services provide a useful deterrent, but care must be taken in assessing the results.

1. Introduction

In the summer of 2010, I received notice that a student of mine had been expelled from the institute where I was employed because of repeated instances of plagiarism. Although there was a considerable amount of supporting evidence for the expulsion, including testimonies from instructors who had taught the student in other courses, it was a report from the internet-based anti-plagiarism service, Turnitin, which sealed the student's fate. Earlier, I had run the student's 2000-word essay through Turnitin's software to search for matches among the billions of web pages as well as the millions of previously submitted assignments in Turnitin's databank. Shortly thereafter, Turnitin's results showed that 38% of the student's text exactly matched several websites. The text of the student's paper was displayed in a column down the left side of the screen with copied sections in various colors corresponding to the web links on the right side where the originals could be found. The student was caught red-handed.

Or so it would seem.

While a matching text rate of 38% appears to be an open and shut case of plagiarism, in fact, there was another student whose Turnitin originality report displayed an even higher percentage. It was only after a closer reading that I realized that the text from the student with the higher figure contained several paragraphs of unquoted copied text which the student had analyzed as part of the term assignment in a course entitled Written and Spoken Discourse. Thus, Turnitin identified a chunk of text as matching which was clearly not a case of plagiarism. In the end, it was confusion over percentages such as this that

* Tel.: +852 2948 8823.
E-mail address: paulstapleton@gmail.com.

1475-1585/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jeap.2011.10.003

triggered the present study, which analyzes the scripts of students writing in their second language (L2) in Hong Kong. In order to better understand the raw percentages generated by Turnitin, as well as to determine what effect this newly introduced tool would have on my students, I ran the essays from two similar classes of graduate-level students through Turnitin's software. However, only one class actually submitted their papers through Turnitin, while the other class had no knowledge of the service.

1.1. Turnitin

As of 2008, 55% of American universities had a license to use an anti-plagiarism product (The 2008 Campus Computing Survey, 2008). Turnitin, the most popular of these services (Gabriel, 2010), claims that it is licensed in 126 countries and available in 10 languages and processes up to 200,000 papers a day (Turnitin, 2010). This compares with a relatively modest 50 countries in 2004 (2004 Fact sheet, 2004). According to iParadigms (2010), Turnitin's parent company, an analysis of tens of millions of papers submitted to Turnitin reveals that regular use of their service produces a 45% reduction of unoriginal content in students' writing after six years of use.

1.2. Intentional and unintentional plagiarism

While the word "plagiarism" is quickly associated with cheating, the issue is clouded not only by whether or not a student intended to steal the words of another, but also by the limits of anti-plagiarism technology and how it is used. In determining whether plagiarism has occurred, the intent of the student writer is a key point. Pecorari (2003) distinguishes between textual and prototypical plagiarism where the former is simply copying without attribution and intent to deceive, while the latter involves deception.

Although the meaning of prototypical plagiarism is clear, textual plagiarism can come in many forms, including the careless omission of quotation marks (or using single instead of double quotation marks) and page numbers (while retaining a citation to the original author). It can also include paraphrasing that shadows the original too closely. Shi (2004, p. 178), for example, identified three levels of textual borrowing in her study:

> (a) with no references, (b) with reference to the author or the source text, and (c) with quotations. Both the first and second categories were further distinguished into three subcategories to indicate whether strings of words borrowed were (a) exactly copied, (b) modified slightly by adding or deleting words or using synonyms for content words, or (c) closely paraphrased by reformulating syntax or changing the wording of the original text.

Keck (2006) offers another variation on this scheme which includes four classifications – Near Copy, Minimal Revision, Moderate Revision and Substantial Revision.

1.3. Plagiarism among L2 students

Further complicating the determination of what defines plagiarism are issues that go beyond the actual mechanics of proper referencing. Howard (2001) coined the term "patchwriting," explained by Pecorari (2003) as a stage that writers new to a certain discourse go through in which their own voice merges with another author's due to a lack of competence, rather than a deliberate attempt to copy. Here, there is no intent to deceive or take credit for another's work; rather, the focus of the composing effort is communicating the message as effectively as possible (Leki & Carson, 1997). Such writing behavior is particularly notable among L2 students who, beyond having to learn academic genres of writing, also have to grapple with the mechanics of a second language. Pennycook (1996) offers an example of this from his interviews with Hong Kong students, whose locale coincides with that of the present study. According to Pennycook, some of his students did not see the point of paraphrasing a piece of text in their L2 that had already been clearly articulated. Another student in his study claimed that her high school teacher valued the ideas expressed over any concerns about the language used to present them. Similar cultural nuances that mitigate, or at least suggest textual, rather than prototypical plagiarism among L2 writers have been observed in other studies as well (Leki, 1992; Spack, 1997).

There is some evidence that plagiarism is more common among L2 students than native speakers of English. Shi (2004), for example, noted in her study comparing the writing of Chinese students with American undergrads whose native tongue was English that the Chinese students borrowed text without attribution at a significantly higher rate than the American students. In the wider East Asian region, Rinnert and Kobayashi (2005), noted in their questionnaire study comparing Japanese and American university students a wide difference between the two in both the amount of training received in properly citing sources, as well as their attitudes toward plagiarism. Specifically, based on results of their survey, American students collectively claimed to have learned much more about paraphrasing and citing correctly compared with their Japanese counterparts. Furthermore, Americans were found to have significantly stronger negative perceptions about plagiarism compared to the Japanese students, over half of whom believed it was conditionally acceptable. These same beliefs (as the Japanese) appear to have existed a considerable length of time ago among students in the same locale as the present study (see Deckert, 1993).

Returning to the dissimilarity between the two basic types of plagiarism, i.e., intentional and non-intentional, the Council of Writing Program Administrators (2008) attempts to further distance the two by claiming that plagiarism and the simple misuse of sources has been unnecessarily conflated.

> Ethical writers make every effort to acknowledge sources fully and appropriately in accordance with the contexts and genres of their writing. A student who attempts (even if clumsily) to identify and credit his or her source, but who misuses a specific citation format or incorrectly uses quotation marks or other forms of identifying material taken from other sources, has not plagiarized. Instead, such a student should be considered to have failed to cite and document sources appropriately.

1.4. Critics of Turnitin

Anti-plagiarism services not only raise ethical issues about a student's degree of intent to copy without acknowledgment, they also raise other concerns. Marsh (2004), for example, laments the socialization of students "toward traditional notions of textual normality and docility [and] a new phase in the bureaucratization of composition instruction." In other words, Marsh views the use of Turnitin as a step toward "conform[ing] to a given set of normalized practices… [that] reify traditional paradigms as progressive writing pedagogy" (p. 436). To address this normalization, Marsh calls for a reexamination of the politics in universities especially in areas where they serve to mold student identity into a prescribed form. The blanket determination of plagiarism also has repercussions for the archives established by services such as Turnitin. Purdy (2009) expresses concern about the integrity of Turnitin's archive which has control over how plagiarism is determined and labeled, which in turn "support[s] ideas of textual ownership and singular authorship privileged by Western culture (p. 73)."

Howard (2007, pp. 11–12) offers perhaps the most powerful statement against electronic anti-plagiarism services in a reaction to Turnitin's claim that its service helps students better understand intellectual property while saving teachers' time.

> [The anti-plagiarism detector] also helps teachers avoid asking the hard questions about what the new revolution in access to text teaches us – that both reading and writing are collaborative, appropriative activities, and that social leaders are not above plagiarism and are not necessarily punished for it. Plagiarism-detecting software also helps teachers describe the issue solely in terms of individual students' ethics, thereby avoiding the difficult task of constructing pedagogy that engages students in the topic and the learning process and that persuades them not just that they will be punished for plagiarizing but that they will [be] able to and glad to do their own writing. In place of the pedagogy that joins teachers and students in the educational enterprise, plagiarism-detecting software offers a machine that will separate them.

1.5. Studies on Turnitin

All of these issues notwithstanding, there remain the more pragmatic concerns over the impact that anti-plagiarism software is having on students' writing behavior. Contrary to the reservations expressed in the theoretical studies reviewed above, the literature shows that the relatively few empirical studies performed in this area are generally supportive of plagiarism-detecting software. For instance, Sutherland-Smith and Carr (2005), in an interview and observational study of teacher's perceptions of Turnitin found a generally positive reaction, although there were concerns that it would not actually save the teachers' time; plus, there were concerns that results would be used to punish students without addressing deeper concerns about plagiarism. Notably, the interviews took place in 2004 and since then, Turnitin has tweaked results to address some of the problems raised. In a study of both teachers and students' perceptions, Savage (2004) found that teachers broadly supported the use of Turnitin as a deterrent to plagiarism; however, they had reservations about whether the originality report caught all instances of it; they also expressed concern about the need to scrutinize all matches. Students were generally supportive as well because, in part, they viewed the service as a way to enable fairer assessment, although concerns were raised related to legal issues and privacy. Similarly, Dahl's questionnaire study of 24 graduate students (2007) found that Turnitin had a formative effect with students being encouraged to seek clarification on what constitutes plagiarism. It also appeared to have a deterrent effect, although two cases of plagiarism were still caught among the group. More recently, in a study of student writing and perceptions among first-year university students in Ireland, Ledwith and Risquez (2008) found that use of Turnitin resulted in a drop in plagiarism with most students holding a positive attitude toward the service. This significant drop in "plagiarism" (actually "matching text") among the scripts of 197 students was noted between two assignments, both of which were screened by Turnitin under the students' full knowledge. The fact that the same group of students submitted essays for both assignments via Turnitin is significant for the present study as we will see below.

Most recently, in a study which most closely mirrors the aims and methodology of the present one, Walker (2010) used Turnitin to track over 1000 scripts from over 500 students at a New Zealand university in two successive assignments, in which students were aware their assignments would be screened for plagiarism. Similar to some of the studies mentioned above (Pecorari, 2003; Shi, 2004), which delineated different forms of plagiarism, Walker screened students' papers for three types of matching test: "sham," "verbatim," and "purloining," where sham most closely mirrors textual plagiarism, verbatim most closely mirrors prototypical plagiarism and purloining is the plagiarizing of another student's writing. One of the criteria

employed in this multifaceted study was a measure of whether the students changed their plagiarism behavior from the first to the second assignment even though both were submitted via Turnitin. Unlike the other studies mentioned here, Turnitin did not register any deterrent effect. However, Walker explained that a sizable number of students who plagiarized on the first assignment, did not submit a second one, suggesting some deterrence; Walker (2010, p. 55) describes this mixed picture:

> However, this [deterrent effect] is difficult to substantiate. Despite the use of Turnitin, there was still evidence of high rates of plagiarism. Although the lecturer fully informed students about Turnitin's capabilities, almost a third of students still submitted a plagiarized first assignment. On being informed in writing by the lecturer that he had detected plagiarism in their first assignment, only a third of those who had plagiarized desisted from further plagiarism in the second assignment. But fully one-fifth of the class either continued to plagiarize or did not expunge existing plagiarism in their assignments.

One further finding from Walker's study of interest to the present one was that international (meaning "L2") students plagiarized significantly more than domestic ones, although fully two-thirds of the L2 students did not plagiarize.

1.6. Research questions

With this backdrop, the present study explores similar issues to that of previous studies using a different research context. While previous studies examining Turnitin's impact have employed interviews, questionnaires and originality reports for one set of students, the present study analyzes Turnitin's results from two distinct, but parallel groups of L2 students, one of which had full knowledge of Turnitin's capabilities and another which was unaware that their essays would be assessed by Turnitin.

The research questions were:

1. Does knowledge that their essays will be screen by the anti-plagiarism service, Turnitin, deter students from plagiarizing?
2. In what ways do the percentages of matching text generated by Turnitin represent plagiarism?
3. Do instances of plagiarism and/or percentages of textual borrowing influence the grades of student writing?

2. Method

2.1. Setting

Due to a fortuitous set of circumstances, the present study could be carried out in order to empirically address the research questions. In March 2010, Turnitin was introduced at the Hong Kong tertiary institute where I work as an optional service for instructors to use in an effort to discourage plagiarism. The timing was fortuitous because I was teaching two very similar classes whose final assignments were due shortly before and after the introduction of Turnitin. This allowed me to receive the final assignments from two distinct, but parallel sets of students: one which was not required to submit their papers via Turnitin, nor had the students any knowledge of the service, and another which was required, and whose students had clear knowledge of the anti-plagiarism software's capabilities.

2.2. Participants and assignment

Participants came from two classes with 22 students each. One was a master's-level class (M.Ed class in TESOL) while the other was at the postgraduate diploma-level (PGD class in English Subject Knowledge). As courses at the latter level can be used to exempt those at the former, the level of the two programs is considered largely equivalent. Both classes were taking the same course, "Written and Spoken Discourse," and the assignment in question required students to analyze a text (preferably education-related) demonstrating mastery of concepts, approaches/theories and techniques of discourse analysis introduced in the course (2000 words). All but one of the students was a native Chinese speaker, and most were either in-service or pre-service teachers.

2.3. Procedure

The M.Ed class submitted their final assignment via email attachment. Although they had received information and warnings about plagiarism from the student handbook and the course instructor, they had not been told anything about Turnitin, and probably were not aware of such a service.

The PGD class had a similar treatment with regard to warnings about plagiarism; however, two weeks before the assignment was due, they were given a half-hour tutorial on how to submit their assignment via Turnitin and shown the software's capabilities in spotting matching text. They were also told they could submit their text to Turnitin as a trial. They were told that if their percentage of matching text was self-assessed as too high, they had the option to revise and resubmit. In summary, the M.Ed class's assignments arrived as email attachments while the PGD class's were received via Turnitin.

After both sets of assignments were received, the data tabulation and analysis began; however, the M.Ed class's assignments first had to be run through Turnitin. Once this was done, there were two sets of results from Turnitin. These were then

checked and adjusted so that they were parallel in all respects. For example, Turnitin has two settings to include or exclude bibliographies and quoted text in its check for matches. These were set to exclude both. There is also an option to set the level at which small matches were excluded. This was set at six words.

Because the assignments were in the area of discourse analysis, some of the texts that the students were analyzing were embedded in their essays without quotation marks and therefore sometimes matched those found on the internet, and thus were included in the students' percentage of matching text in Turnitin's originality report. These were excluded from the data analysis, but this exercise in itself foreshadowed the need to look closely at results generated from Turnitin.

Turnitin generates textual matches between the student essay and an Internet website, or texts in a databank of past student assignments, and presents these as chunks of text side-by-side in corresponding colored fonts along with a percentage of the total essay for each matching chunk. In order to determine a final score of plagiarism, I examined the matching chunks of text written by each student for whether the matching text was 1) a case of textual plagiarism, i.e., no deception intended; 2) probable coincidental matches or near copies; or 3) prototypical plagiarism, i.e., deception was intended. These criteria require further explanation.

In the present study, Turnitin's default setting of six or more successive matching words was employed. In the case of #1 above, if a citation was included, but no quotations or page numbers appeared, the isolated text was deemed textual plagiarism, i.e., a case of poor citation practice. Here, a number of variations appeared. Apart from the simple omission of quotations or page numbers (see (i) below), there were also cases where attribution to a certain idea or concept appeared (without quotations or page numbers), but was only implied through proximity to the copied text (see (ii) below). All such cases where a transparent attribution to the original was included, but referencing norms for quotations were not followed were considered instances of textual plagiarism. Two examples of textual plagiarism taken from the corpus of students' assignments are below:

i) five types of general functions performed by speech acts: declarations, representatives, expressives, directives and commissives (Yule, 1996) (No quotations were used; taken from Yule, 1996, p. 53).
ii) Nunan (1991) states that [skilled writers will revise their writing at all levels of lexis, sentence and discourse so writing class should not be just concern (sic) with the mechanics of grammar, spelling, punctuation and vocabulary.] (The section inside the square brackets was copied; underlined words added by the student; taken from Nunan, 1991, p. 90).

Cases of #2 above, probable coincidental matches or near copies, consisted of strings of between six and ten successive words without attribution.[1] Most of these included common stock phrases, which were isolated as matching text by Turnitin. Within this category were instances where student authors occasionally borrowed heavily from an existing text. This patchwriting (see example in the Discussion), in which an author copies a text and then replaces a certain number of words within was sometimes revealed by Turnitin. For the purposes of this study, when cases of such patchwriting appeared, so long as the strings of text did not exceed ten words, they were deemed near copies. Indeed, determining whether a string of matching text was coincidental also demanded close scrutiny because even eleven or more consecutive words can match coincidentally.[2] This category also included matches that fell outside any of the other two categories, e.g., web links and titles.

Those matches that could not be classified either in #1 or #2 were deemed prototypical plagiarism, i.e., #3. The following sample taken from the students' assignments is clearly prototypical plagiarism because the citations themselves have been copied and one word (underlined) replaces the original:

> Most research findings point out that explicit teacher feedback can play a positive role. (Ferris & Roberts, 2001; Hyland, 1998). In the second language context, teachers' specific and quality comments can lead to substantial student revisions that improve the quality of writing. (taken from the original, Li, 2007)

3. Results

In Table 1, the percentages generated by Turnitin for each student paper are shown (after the embedded texts for analysis, mentioned above, had been eliminated). For the M.Ed class, matching text ranged from a high of 44 percent to a low of one percent with a mean percentage of 12.1 percent (SD 11.3). The PGD class ranged from 12 percent to zero with a mean percentage of 4.1 (SD 3.1). While these raw figures do provide some indication of differences in plagiarism between the two classes, they remain only rough. Thus, Table 1 also indicates the levels of textual and prototypical plagiarism as well as coincidental matches and near copies in the two classes. For the M.Ed class, textual plagiarism ranged from a high of 13 percent to a low of zero with a mean of 3.6 percent (SD 3.8) while prototypical plagiarism ranged from 29 percent to zero with a mean of 4.8 percent (SD 8.1). For the PGD class, textual plagiarism ranged from seven percent to zero with a mean of 2.2 percent (SD 2.3), while prototypical plagiarism ranged from two percent to zero with a mean of less than one percent. When a Student's *t*-test (2-tailed) was performed on these three categories between the two groups, for the matching text

[1] The number "ten" was chosen because it is near the outer limited of what one can reasonably call a coincidental match.

[2] It is interesting to note that the following ten-word string generates 151,000 exact matches in an advanced Google search: "the purpose of this study is to explore whether the."

Table 1
Means and (standard deviations) of percentages of different types of matching text.

Category	M.Ed ($n = 22$)	PGD ($n = 22$)	Significance level p
Matching text (from Turnitin)	12.1 (11.3)[a]	4.1 (3.1)[a]	.01
Textual plagiarism	3.6 (3.8)	2.2 (2.3)	.14
Prototypical plagiarism	4.8 (8.1)	.2 (.7)	.05
Coincidental matches and near copy	3.9 (4.0)	1.5 (1.0)	.01

[a] The totals for textual and prototypical plagiarism plus coincidental matches do not equal "matching text" (as they should) because Turnitin generates percentages to the nearest round figure and includes many instances of "<1%."

percentages generated by Turnitin and prototypical plagiarism, significant differences were found, $p \leq .01$ and $p \leq .05$ respectively; however, no statistical difference was found in the rate of textual plagiarism. As for the amount of text that was categorized as coincidental matches or near copies, the mean percentage for the M.Ed class was 3.9 (SD 4.0) which was significantly greater ($p \leq .01$) than the mean for the PGD class which was 1.5 (SD 1.0).

4. Discussion

While the present study is quantitative in nature, the interpretation of the numbers generated after the raw scores were received from Turnitin, i.e., making decisions about whether or not a chunk of text was intentional plagiarism, or even a coincidental match, was somewhat qualitative. Accordingly, in this section, both the numerical data as well as the qualitative issues arising from the analysis will be discussed. However, before proceeding, some discussion about intentionality, the notion at the heart of this study, appears appropriate.

The intentions of the student authors in this study can be uncovered to a certain extent based on matching text evidence. However, such evidence can only take us a certain distance. It is not proof positive of intention. For example, if a student author places copied text in his or her paper without quotations while including an accurate attributive citation, one can reasonably assume that it was non-intentional plagiarism, i.e., simply poor citation practice. However, if a student copies a string of text without quotations while including a citation to an author unconnected to the quotation, presumably as a distracter, it can reasonably be deduced as a case of intentional deception. Intentionality, however, cannot be proved beyond doubt in either of these cases. Even interviews with students in which they are asked about their writing behavior cannot necessarily get closer to their true intentions, especially in the case of prototypical plagiarism. Interview data from a recent study on plagiarism among students in the same locale as the present study attest to this (Li & Casanave, forthcoming). Accordingly, the findings in the present study are simply approximations based on the best evidence at hand, the text itself.

4.1. Statistical data (question 1)

The first research question inquired whether students would be deterred from plagiarizing if they knew that their essays would be screened by anti-plagiarism software. Although the answer to such a question may appear self-evident, the findings provided by this study shed light on various nuances that accompany the issue. The group of students who had knowledge and hands-on experience with Turnitin (the PGD class) appeared to be deterred from plagiarizing when compared with the class that had no such knowledge (M.Ed class). In fact, many in the PGD group did submit their essays to a dummy site first as a trial before their final submission, a fact which underscored their familiarity with notions of plagiarism. This deterrent effect is in line with the claims made by Turnitin as well as other studies (Dahl, 2007; Ledwith & Risquez, 2008; Savage, 2004), but somewhat contrary to findings from Walker (2010). However, it should be noted here that in Walker's study, all of his students submitted both their first and second assignments via Turnitin, which is unlike the present study where half of the students were completely unaware that their assignments were being assessed by the anti-plagiarism software. A closer look at the percentages reveals that among the PGD class, the highest percentage of prototypical plagiarism, i.e., the intentional kind, of any student was only two percent, or about 40 words in a 2000-word essay. In the M.Ed class, however, nine students were assessed as plagiarizing a greater amount than this, with two students plagiarizing over 25 percent of their text, and six more having copied over five percent. On the other hand, among the PGD and M.Ed students there were 18 (82%) and 11 (50%) students respectively who registered no instances of prototypical plagiarism at all, a result not dissimilar to Walker (2010) who found that roughly three-quarters of the scripts in his study had no plagiarism at all. In other words, based on the limited example of these two groups, it can be said that for a sizable number of the students, no deterrent was necessary, i.e., Turnitin probably served as a deterrent only for those with a propensity to plagiarize. The high standard deviation scores for the M.Ed class bear this out with a few, especially two students, skewing the data. Fig. 1 shows this graphically.

The one category where no significant difference was found was textual plagiarism. Essentially, students in both classes unintentionally plagiarized at a similar rate, with fewer than half of the students in each class accounting for virtually all of these cases. As textual plagiarism can be put down to either carelessness or ignorance, it lacks intention. Thus, one would expect rates of these two traits to be similar.

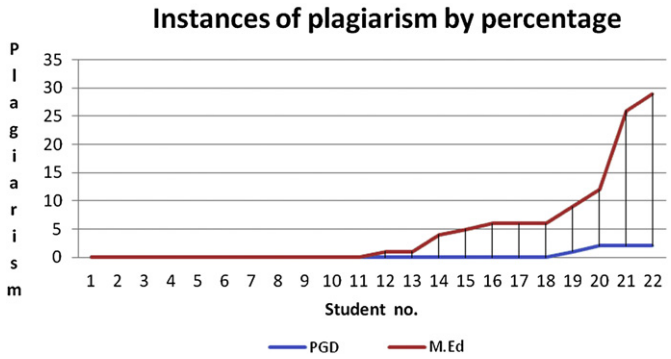

Fig. 1. Instances of plagiarism by percentage.

As for coincidental matches and near copies, the significant difference between the two groups suggests that something more than coincidence was occurring, particularly because it was the M.Ed group that scored a higher percentage of instances. Once again, the high score for this group can be attributed to a few individuals who copied and then manipulated chunks of text. This may be a common strategy used by L2 students when they write essays as illustrated in a recent study (Stapleton, 2010) in which an individual L2 student employed a system in which chunks of text were copied and pasted into an essay and keywords were replaced in an effort to make the sentences her own, as well as to avoid plagiarism. An example of the sort of near copy tagged as a match by Turnitin appears below with copied text in bold and the original in parentheses.

> Pedagogic discourse cannot reflect the (Such pedagogic discourse is not reflective of the) **concerns in the integrative curriculum education system**. (The) **Teaching of** English language in (Malay Language subject in the) **the integrative curriculum** needs (demands) **the incorporation of thinking skills among students**. **To** develop logical, rational, analytical and objective (acquire thinking skills that incorporate logic, rationality, analytical skills and objectivity) **thinking skills**, classroom **teaching and learning activities** must be prepared (have to be geared) **to** encourage **students**.

As for the remaining category, "matching text," i.e., the percentage generated by Turnitin, while somewhat crude when compared to the more specific results provided by the other categories, this raw score does indeed appear to be a strong indicator revealing a statistically significant difference between the two classes, and as we have seen, this is supported by the more detailed categories. However, it is the differences between the raw figures provided by Turnitin and those that were generated by a more in-depth examination of the matching texts where qualitative issues emerge.

4.2. Qualitative issues (question 2)

Perhaps the most notable qualitative issue arising from this study is the complexity of assigning a numerical value to a chunk of text that has its own intrinsic properties steeped in the cognitive processes of its author. As software can only generate the numerical extent to which two chunks of text match, it fails to make allowances for the intentions (Pecorari, 2003) or cultural background (Pennycook, 1996) of the student author. As this study shows, textual, or unintentional plagiarism constituted a significant portion of the matching text, as did small, coincidental matches and near copies. In the case of the PGD class, almost all of the matching text was in fact non-intentional plagiarism. Thus, there is an inherent danger in relying on the Turnitin's numbers alone. In fairness to Turnitin, it should be noted that the company goes to considerable lengths to distance themselves from claiming that their originality report on a student's essay is a proof positive indicator of plagiarism (Turnitin Instructor User Manual, 2010, p. 48).

> Warning: These indices do not reflect Turnitin's assessment of whether a paper has or has not been plagiarized. Originality Reports are simply a tool to help an instructor find sources that contain text similar to submitted papers. The decision to deem any work plagiarized must be made carefully, and only after in-depth examination of both the submitted paper and suspect sources in accordance with the standards of the class and institution where the paper was submitted.

It is at this point that this study's second research question, namely, *In what ways do the percentages of matching text generated by Turnitin represent plagiarism?* can be addressed. Certainly, a study such as this one would have been impossible a decade ago and would still be very difficult without the type of software provided by Turnitin. In this sense, Turnitin offers an exceedingly convenient way to isolate matching chunks of text. Additionally, the percentages generated for each paper do

tend to provide an overall indication of whether a student has lifted material from another source. On the other hand, as the statistics for textual plagiarism, coincidental matches and near copies in Table 1 indicate, a rather large portion of matching text may be unintentional, and it is here that Turnitin's caveat above needs underscoring.

4.3. Student grades (question 3)

As for the third question which inquired whether instances of plagiarism and/or percentages of textual borrowing influence the grades of student writing, two noteworthy points arose. The first of these was a comparison of grades across the two classes. Since the essays from both classes were graded (by the author) and moderated by another grader according to criteria-referenced rubrics before the present study was carried out, it is safe to conclude that any difference between the grades in the two classes is incidental, rather than being related to the purposes of this study. In any event, both classes scored a comparable number of A's, B's, C's and F's. In other words, the differences in the degree of plagiarism noted by this study did not appear to have any effect on the collective grades. One reason for this rather surprising finding is that, among the M.Ed students, the student with the most egregious instance of plagiarism (who was expelled, as discussed at the beginning) was not included as she received no grade. And, regrettably, the second most egregious case of intentional plagiarism was simply not noticed, which again underscores the efficacy of Turnitin which turned up this case of prototypical plagiarism, sadly, too late for punishment. Secondly, when a cross-check was performed to determine whether there was any correlation between the grades received by individual students and their extent of prototypical plagiarism, no pattern was noted, except that the students receiving grades in the top 15 percent of the class tended to have very few instances of unquoted matching text. Such findings, if further replicated, may be worth repeating to students before essays are assigned. To put it succinctly, plagiarizing appeared unlikely to produce a higher mark.

4.4. Reflecting on criticisms of anti-plagiarism software

Returning to Howard's criticism of Turnitin (2007) in which she claims that anti-plagiarism services create a chasm between teachers and students which ends up hindering discussions about engagement with texts, there may be an alternative perspective. Given that despite severe penalties, plagiarism has persisted on campuses for generations, it can be argued that anti-plagiarism technology is actually shedding new light on an old, but enduring issue and stimulating new discussions about what constitutes plagiarism. Instead of repeating the tired threats about the evils of plagiarism that are frequently directed toward students, services like Turnitin are bringing to the fore the need to precisely define how student writers should interpret, summarize, paraphrase or quote, in other words, engage with texts. Rather than being viewed as a tool of bureaucratization (Marsh, 2004) or another step toward westernizing the culture of writing (Purdy, 2009), Turnitin can be seen as an opportunity to broaden the discussion. After all, the results that Turnitin generates are simply numbers, and all numbers need interpretation. And as this study has shown, the numerical product is simply a starting point for instructors' interpretation.

Such a comment may appear daunting. Indeed, what teacher needs yet another task to further complicate the grading process? The reality, however, unlike the bureaucratization envisioned by Marsh (2004), is probably much simpler than appearances would suggest. Among the group of students who were aware that their papers were being screened by Turnitin (the PGD class), only two papers generated a raw percentage from Turnitin in excess of seven percent. In the present study, such a low percentage (seven percent and below) turned up no egregious examples of plagiarism. In most cases, matching text of seven percent or less was an accumulation of bits and pieces of coincidental matches, near copies and, when it was plagiarism, it was easily spotted in the colored text. Finding these took minimal amounts of time.

5. Conclusion

The deterrent effect provided by Turnitin in this study certainly comes as no surprise. The degree to which plagiarism failed to appear in the PGD class was possibly because students were afraid of being caught by Turnitin, and this aligns with the anti-plagiarism detector's own statements, as well as those from several other studies about the service's role as a deterrent. Therefore, if one agrees that the copying of another's words without attribution is to be avoided, Turnitin is providing a useful service. However, as illustrated by this study, care must be taken in interpreting the originality report. Accordingly, university administrators introducing Turnitin or other anti-plagiarism services need to ensure that all users, both students and instructors, understand that matching text percentages are simply rough indicators, rather than definitive conclusions about the extent of plagiarism within a text. Accompanying this discussion, clearer guidelines need to be disseminated regarding what exactly constitutes plagiarism and how it can be avoided. Such a discussion is important in order to avoid Turnitin becoming a tool for student manipulation, as opposed to a useful instrument for internalizing academic values (Ledwith & Risquez, 2008).

An important point remains regarding the native language of this study's participants. All except one student were non-native English speakers, which may have significant implications. As noted earlier, there is considerable evidence suggesting that L2 writers plagiarize at a higher rate than their native speaking counterparts (Rinnert & Kobayashi, 2005; Shi, 2004; Walker, 2010). These findings suggest that extra attention and instruction in plagiarism avoidance may be necessary for students whose native tongue is not English or those who have been educated outside the Western world. Encouragingly, the

findings of the present study indicate that the Chinese students may have reduced their tendencies to plagiarize when presented with an instrument (Turnitin) that enhanced their awareness of potential problems with their writing. This new tool, coupled with classroom instruction that addresses citation conventions (Shi, 2004) has the potential to bring L2 writers closer to mainstream practices, however imperfect they may be. It should be repeated, however, that similar to Walker's cohort of L2 students, a majority of students in both groups did not intentionally plagiarize.

As is customary in small-scale research, it should be noted that the findings of the present study were based on a small number of students and cannot be generalized to larger populations. Because the study was not designed in a way that students in each group were randomly assigned, it is possible that the students' individual histories influenced the results; thus, the findings herein are indicative only. However, as anti-plagiarism software continues to increase its presence in scholarly affairs, clearly, more studies are needed. Questions such as: 1) in what way does this technology change the paraphrasing skills of students? 2) what sections of essays are more likely to contain matching text? or 3) how does the technology alter the writing process?, all remain unanswered.

This study of Turnitin results has brought to the fore the need for transparency, both in defining plagiarism, and in providing students the autonomy to use Turnitin to check their papers for matching text before submission (Ledwith & Risquez, 2008). It is noteworthy that the PGD class was able to trial their submissions in Turnitin before sending in their final papers. Such an option opens up the possibility of the manipulation of wording by students simply in order to avoid being caught by Turnitin, which most instructors would find an unacceptable engagement with text. This shows that anti-plagiarism services, like all new technologies should be greeted with caution and judged on their merits. However, the advantage of having an instrument, such as Turnitin, which if used effectively, encourages a clearer definition of plagiarism, appears to outweigh the former non-technological one in which plagiarism was ill-defined and often difficult to prove.

References

Council of Writing Program Administrators. (2008). *Defining and avoiding plagiarism: The WPA statement on best practices*. Retrieved from. http://wpacouncil.org/positions/WPAplagiarism.pdf.

Dahl, S. (2007). TurnitinR: the student perspective on using plagiarism detection software. *Active Learning in Higher Education, 8*, 173–191.

Deckert, G. D. (1993). Perspectives on plagiarism from ESL students in Hong Kong. *Journal of Second Language Writing, 2*, 131–148.

2004 Fact sheet. (2004). *Turnitin*. Retrieved from. http://www.Turnitin.com/static/press/factsheet2004.pdf.

Gabriel, T. (2010 July 5). To stop cheats, colleges learn their trickery. *New York Times*, Retrieved from. http://www.nytimes.com/2010/07/06/education/06cheat.html?_r=2&adxnnl=1&adxnnlx=1278453675-7ZgyKd2w+enFiStBdjOw6Q.

Howard, R. M. (2001). Plagiarism: What should a teacher do? Paper presented at the Conference on College Composition and Communication, March 17, in Denver, USA.

Howard, R. M. (2007). Understanding "internet plagiarism". *Computers and Composition, 24*, 3–15.

iParadigms. (2010). *A summary of the effectiveness of Turnitin*. Retrieved from. http://Turnitin.com/resources/documentation/Turnitin/company/Turnitin_Summary_of_Effectiveness_hires.pdf.

Keck, C. (2006). The use of paraphrase in summary writing: a comparison of L1 and L2 writers. *Journal of Second Language Writing, 15*, 261–278.

Ledwith, A., & Risquez, A. (2008). Using anti-plagiarism software to promote academic honesty in the context of peer reviewed assignments. *Studies in Higher Education, 33*, 371–384.

Leki, I. (1992). *Understanding ESL writers: A guide for teachers*. Portsmouth, NH: Heinemann.

Leki, I., & Carson, J. (1997). Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly, 31*, 39–69.

Li, X. (2007). Identities and beliefs in ESL writing: from product to processes. *TESL Canada Journal, 25*, 41–64.

Li, Y., & Casanave, C. P. (forthcoming). Two first year students' strategies for writing from sources: patchwriting or plagiarism? *Journal of Second Language Writing*.

Marsh, B. (2004). Turnitin.com and the scriptural enterprise of plagiarism detection. *Computers and Composition, 21*, 427–438.

Nunan, D. (1991). *Language teaching methodology: A textbook for teachers*. New York: Prentice Hall.

Pecorari, D. (2003). Good and original: plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing, 12*, 317–345.

Pennycook, A. (1996). Borrowing others' words: text, ownership, memory and plagiarism. *TESOL Quarterly, 30*, 201–230.

Purdy, J. P. (2009). Anxiety and the archive: understanding plagiarism detection. *Computers and Composition, 26*, 65–77.

Rinnert, C., & Kobayashi, H. (2005). Borrowing words and ideas: insights from Japanese L1 writers. *Journal of Asian Pacific Communication, 1*, 31–56.

Savage, S. (2004). Staff and student responses to a trial of Turnitin plagiarism detection software. In *Proceedings of the Australian Universities Quality Forum 2004*, Retrieved from. http://www.auqa.edu.au/auqf/2004/program/papers/Savage.pdf.

Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication, 21*, 171–200.

Spack, R. (1997). The acquisition of academic literacy in a second language: a longitudinal case study. *Written Communication, 14*, 3–62.

Stapleton, P. (2010). Writing in an electronic age: a case study of L2 composing processes. *Journal of English for Academic Purposes, 9*, 295–307.

Sutherland-Smith, W., & Carr, D. (2005). Turnitin.com: teachers' perspectives of anti-plagiarism software in raising issues of educational integrity. *Journal of University Teaching and Learning Practice, 3*(1b), 94–101.

The 2008 Campus Computing Survey. (2008). *Colleges invest in emergency notification*. Retrieved from. http://www.campuscomputing.net/survey-summary/2008-campus-computing-survey.

Turnitin. (2010). *Quick facts*. Retrieved from. http://turnitin.com/static/resources/documentation/turnitin/sales/Turnitin_Quick_Facts.pdf.

Turnitin Instructor User Manual. (2010). *Chapter 2: Originality checking*. Retrieved from. https://www.turnitin.com/resources/documentation/turnitin/training/en_us/Instructor_Manual_en_us.pdf.

Walker, J. (2010). Measuring plagiarism: researching what students do, not what they say they do. *Studies in Higher Education, 35*, 41–59.

Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press.