

Patterns in plagiarism and patchwriting in science and engineering graduate students' research proposals

Michelle Vieyra

Department of Biology and Geology,
University of South Carolina Aiken, USA

michellev@usca.edu

Denise Strickland

Center on Education Policy and Workforce Competitiveness,
University of Virginia, USA

DCStrickland@virginia.edu

Briana Timmerman

South Carolina Department of Education, USA

briana.timmerman@gmail.com

Keywords: plagiarism, patchwriting, graduate students, science disciplines, STEM

Abstract

As part of a larger study, written research proposals were collected from 115 science and engineering master's and doctoral students and reviewed by *SafeAssign*TM with approximately one-third of them containing sentences that were plagiarised as previously reported in Gilmore, Strickland, Timmerman, Maher and Feldon (2010). (We use the term plagiarism, but do not imply any intentional deceit by the students.) Here we report on the patterns of plagiarised material in the hope that it will contribute to the growing awareness of the problem of plagiarism in graduate schools as well as provide insight into the causes of plagiarism. Instances of plagiarism were coded as to 1) the type of source material (primary, secondary, technical, or popular literature), 2) the nature of the inappropriate use (directly copied, a few words changed, minor grammar alterations, or attempted but insufficient paraphrasing), 3) where in the proposal (introduction, methods, results, or discussion) the plagiarism appeared, and 4) whether or not the plagiarised information was cited and if it was, whether or not the citation was accurate. Plagiarised text was found in 28% of the proposals. Clustering of certain patterns of behaviour, such as directly copying material from popular literature while paraphrasing information from primary scientific literature, were examined in an attempt to gain insight into the cause of the plagiarism. It is our interpretation that the source of the plagiarism was a lack of familiarity with scientific writing as a genre and lack of awareness of its norms and conventions.

Introduction

Plagiarism in the academic setting has been defined as literary theft, intellectual theft or academic dishonesty and includes copying another's words or ideas without appropriately crediting the author (Eret & Gokmenoglu, 2010). It has been cited as the most common form of academic misconduct at universities (Breen & Maassen, 2005; McCabe, 2005; Selwyn, 2008) and it has been examined among

undergraduate students by many researchers (Breen & Maassen, 2005; DeVoss & Rosati, 2002; Franklyn-Stokes & Newstead, 1995; McCabe, Trevino, & Butterfield, 2001; Pickard, 2006; Selwyn, 2008).

Many studies of plagiarism among undergraduates focus on first year students and the assumption may be that plagiarism issues disappear by graduation as students learn how to appropriately use and cite resources. A survey of master's students in an education programme (Love & Simmons, 1997) revealed that these graduate students were not given any training on proper citation. Further, they felt that instructors assumed that since they were graduate students they should already know how to avoid plagiarism (Love & Simmons, 1997). Perhaps due to this assumption, studies of plagiarism among graduate students have been relatively limited as compared to studies of undergraduate plagiarism. The assumption that graduate students should already be proficient at reading and citing primary literature may have had more merit when the diversity of student background was more limited. In today's international educational arenas however, students enter graduate schools across the globe with a wide array of backgrounds and cultural contexts and making such an assumption appears to be a grave disservice. McCullough and Holmberg (2005) examined the prevalence of directly copying and pasting text from a source by doing a Google search of phrases from master's theses. Twenty-seven percent of the master's theses examined had at least one instance of plagiarism. Sources used ranged from popular websites to primary journal articles. The highest incidence of plagiarism was found in science and engineering theses. Plagiarism has also been noted in personal statements submitted for medical residency programmes (Cole, 2007; Segal et al., 2010). Segal et al. (2010) assessed just under 5,000 residency applications and found significant examples of plagiarism in 5.2% of them.

A preliminary survey of 115 research proposals submitted by science and engineering graduate students at three US universities revealed a high incidence of plagiarism (Gilmore, Strickland et al., 2010). Rates were similar for master's and PhD students across several science and engineering disciplines. Plagiarism rates were lower for students who spoke English as a first language. The purpose of this study is to further characterise instances of plagiarism found in the graduate student research proposals surveyed by Gilmore et al. (2010). Four areas of particular interest are:

1. the specific **types of plagiarism** observed (directly copying source material or copying and then making slight changes to source material);
2. the **types of sources** from which the plagiarised text was derived (popular website, technical website, secondary journal articles, or primary journal articles);
3. the **citation status** of the text in question (no citation, incorrectly cited or correctly cited);
4. the **location** of citation errors within the student's paper (introduction, methods, results or discussion).

Identifying these four aspects of plagiarised text is a necessary precursor to then delineating patterns in the interplay between these factors.

Review of the literature

Students may resort to acts of plagiarism for a variety of reasons including an overall lack of understanding of how to correctly cite the sources they use and the pressure to complete work (Breen & Maassen, 2005; DeVoss & Rosati, 2002; Eret & Gokmenoglu, 2010; Jensen, Arnett, Feldman, & Cauffman, 2002). A study by Breen and Maassen (2005) showed that although students are concerned about avoiding plagiarism they often do not understand what it means to paraphrase or when it is appropriate to cite information.

There is no doubt that source use and citation are important in the sciences, serving a variety of rhetorical purposes (Harwood, 2009; Salager-Meyer, 1999). Expert scientific writers use citation differently than novice writers (Mansourizadeh & Ahmad, 2011). Petrić (2007) found that writers of high-rated master's theses used citation for a greater variety of purposes than writers of low-rated master's theses.

However, there is confusion within science disciplines about what is and is not acceptable (Couzin-Frankel & Grom, 2009; Roig, 2009). This confusion is then passed on to students. Research shows that professors disagree about what is and is not plagiarism (Roig, 2001), and about whether it is an ethical or developmental issue (Abasi & Graves, 2008; Crocker & Shaw, 2002; Sutherland-Smith, 2005). This creates a dilemma for students in which they may want to use source texts appropriately, but lack appropriate strategies and knowledge (Flowerdew & Li, 2007; Pecorari, 2003; Roig, 1999).

Studies show that undergraduate science and engineering courses may have the highest occurrence of plagiarism (Simon et al., 2004; Sheard, Dick, Markham, McDonald & Walsh, 2002; Yeo, 2007), potentially because the information found in primary sources is often couched in mathematical formulae and scientific jargon that is difficult for students to understand and paraphrase (Roig, 1999). Additionally, the approach to gathering information and working through tasks is fundamentally different when compared to other disciplines (Yeo, 2007).

This confusion, however, does not protect students from accusations of plagiarism. Sutherland-Smith (2005) found that 2 out of 11 writing instructors they interviewed automatically assume that inappropriate text reuse is intentional and therefore plagiarism. The honour codes of all three universities from which this sample was taken, if applied to the writers of the texts we analysed, would deem them as plagiarists, with potentially severe consequences.

For the purposes of this study, we use the term plagiarism to indicate "textual plagiarism", a term defined by Pecorari (2008, p. 4), as "the use of words and/or ideas from another source, without appropriate attribution". Textual plagiarism may or may not involve the intention to deceive, and unintentional plagiarism may be manifest in the form of patchwriting, the textual 'pastiche' of copied and original text which novice writers produce as they develop as independent writers (Howard, 1999). In this study, we seek to explore the relationships between students' writing strategies and other aspects of their text reuse, such as source type and use of citation.

Methods

Study context

This research was conducted as part of a larger three-year study, beginning in the 2007/2008 academic year, examining science and engineering graduate student research skill development. Participants were drawn from two universities in the Southeastern United States and one university in the Northeastern United States. As part of this larger study, research proposals were collected from graduate students. Participants were given instructions on how to develop their research proposals including descriptions of what sections to include (introduction, methods, anticipated results and discussion). Students were asked to use citation styles typical of their fields but no instructions were given regarding proper citation or appropriate sources. In addition, demographic information was collected including gender, field of study, degree sought (PhD or master's), year of study, and English fluency (based on self-reported English as a Second Language (ESL) status and Test of English as a Foreign Language scores (TOEFL)). At these institutions, TOEFL scores indicated that the student was a non-native English speaker and international (not a US citizen). See Table 1 for participant demographic information.

Table 1:
Demographics of participants who plagiarised/total participants

		Number (%)
Gender	Male	20/65 (31%)
	Female	12/50 (24%)
Degree Sought	Masters	13/47 (28%)
	Doctorate	19/68 (28%)
English 1st Language	Yes	11/70 (16%)
	No	21/45 (47%)
Discipline	Biological Sciences	8/29 (28%)
	Engineering	4/24 (17%)
	Marine Science	4/17 (24%)
	Math/Statistics	3/11 (27%)
	Health Sciences	5/9 (56%)
	Technology	4/8 (50%)
	Chemistry	3/6 (50%)
	Science Education	1/5 (20%)
	Geological Sciences	0/5 (0%)
	Physics	0/1 (0%)
Total		32/115 (28%)

Text analysis tool

Samples were submitted to *SafeAssign™*, a service that finds both exact and inexact matches between submitted work and text that is available online or through certain databases (Blackboard Inc., 2010). *SafeAssign™* is not a perfect tool and will not detect matches to all possible source materials (such as printed books or journals that have not agreed to be part of the *SafeAssign™* database); therefore our results represent a conservative estimation of the extent of plagiarism in these papers. One benefit of using a program like *SafeAssign™* however, is the ability to evaluate a relatively large sample of papers. This programme also provides access to the original source material for the matched text, allowing for a detailed analysis of the extent and type of text matching observed. Only sentences flagged by *SafeAssign™* were evaluated in this study.

Coding scheme development and unit of evaluation

Two investigators independently evaluated each *SafeAssign™* report which evidenced inappropriate text appropriation. A coding scheme to describe the nature of the plagiarised text was developed through the constant comparative approach (Glaser, 1965). Disagreements between raters were resolved via discussion until consensus was reached (Johnson, Penny, Gordon, Shumate, & Fisher, 2005). Instances where *SafeAssign™* flagged a sentence that, upon evaluation, was not plagiarised were discussed by the evaluators and excluded from further analysis. Typically these were short headings or adequately paraphrased sentences.

We found it appropriate to evaluate plagiarism at the sentence level rather than by passage or instance because even when students use the same source across an entire paragraph they may directly copy some sentences yet paraphrase others, cite some sentences and not others, merge sentences or separate sentences with material from other sources. In addition, they may have instances where one or two sentences are copied versus instances where entire paragraphs are copied. Evaluation at the sentence level allowed for differentiation of the relative extent of plagiarism across the whole paper as well as a deeper analysis of each sentence across several dimensions. For each problematic sentence, the type of plagiarism present, the type of source material from which the text was appropriated, whether or not the student cited the source of the sentence and its location in the student's paper

were evaluated. Problematic sentences were coded using specific categories and criteria discussed below.

Coding scheme for type of plagiarism

For each sentence containing problematic text, the type of plagiarism was recorded as either Direct Copying, Word Change, Grammar Change or Complex (see Table 2 for descriptions and Table 3 for examples). These codes are not mutually exclusive and, in fact, codes indicating greater distance from the source text may include behaviours described by codes more proximal to the text. Sentences were assigned the most distal relevant code. For example, a sentence could be 90% direct copied with some word changes and a verb tense change and would be coded as 'Grammar Change'.

Table 2:
Data coding scheme for type of plagiarism

Type of Plagiarism	Description	Code used in figures
Direct Copy	Verbatim copying.	DC
Word Change	Nearly verbatim copying with a few words replaced by synonyms.	WC
Grammar Change	Whole sentence fragments copied verbatim, but writer reorganises the order in which they appear in the sentence and/or changes verb tenses.	GC
Complex	Writer attempts to paraphrase by using multiple techniques listed above, but much of the sentence is still recognisable as copied and/or the material is not cited.	CX

Table 3:
Examples of the types of plagiarism coded

Type of Plagiarism	Original Source	Student Text
Word Change	Although SST above 28°C is not sufficient to produce organized deep convection in the tropical atmosphere, it is a necessary condition.	Although having SST > 28 is not sufficient to produce organized deep convection in the tropical atmosphere, it is a necessary condition.
Grammar Change	In 1987, the Southern Regional Education Board (SREB) and 13 Southern states formed the SREB-State Career/Technical Education Consortium and established two goals.	In 1987, the Southern Regional Education Board (SREB) and the 13-state State Vocational Education Consortium established two goals.
Complex	Each monomer is characterized by <ul style="list-style-type: none"> • intracellular N-terminal domain of 61 amino acids • single pass transmembrane domain of 28 amino acids • large extracellular C-terminal domain of 671 amino acids which is soluble and contains the Tf binding site and a trypsin-sensitive site 	Each monomer contains a large extracellular C-terminal domain that contains the Tf-binding site, a single-pass transmembrane domain and a short intracellular N-terminal domain. (<i>NOT CIT-ED</i>)

Coding scheme for type of source materials

The origins of problematic text were identified as primary sources, primary source abstracts, secondary sources, popular technical works, or popular works. A Primary Source, in science and engineering fields, is a published work which presents original research conducted by the authors and includes a methods section and actual data. The source was coded as Primary Abstract when the student only used text from the abstract. Secondary Sources, in science and engineering fields, are published works which review the research of others. Examples include textbooks or journal review articles. Popular Technical works include source material available on the internet, but intended for a specialised or technical reader (i.e. drug information sheets, lab protocols, patents, etc.). Popular works include source material which is published on a website and intended for a general audience.

Citation status

Norms and conventions of appropriate (non-plagiarised) writing in science and engineering fields focus on issues of citation. Citations are required for all claims of factual knowledge (knowledge creation) so that findings or interpretations later found to be in error can be tracked through the literature and nullified. Consequently, our coding scheme differs from others in the patchwriting field (e.g. Howard, Serviss, & Rodrigue, 2010) in that whether or not appropriated text is properly cited is critical. Problematic text flagged by SafeAssign™ was coded as to whether it was cited and if so, whether or not that citation was correct. Care was taken by the evaluators to assure that sentences coded as not cited were truly not cited anywhere in the paper (whether after the sentence being evaluated or at the end of the paragraph or page). When students cite information from review articles or the introduction of primary articles, additional types of citation errors can occur (as noted by Pecorari, 2006). To determine whether the appropriate source was cited, evaluators consulted the original source material used and compared it to the text and citations within the student paper.

Location of plagiarised text

Plagiarised sentences were coded as occurring in the Abstract, Introduction, Methods, Results or Discussion sections of the student's proposal.

Results

Overall rates of plagiarism

Of the 115 papers analysed, 32 (28%) had at least one sentence with inappropriate text re-use. Across the sample there were 380 plagiarised sentences. Extent of plagiarism ranged from insufficiently paraphrasing a single sentence in the paper to up to 64% of the paper copied directly from an internet source with none of the sources cited.

Types of plagiarism

Direct Copying of text was the most common type of plagiarism detected (see Table 2 for code definitions). Specifically, 68% of the plagiarised sentences were Directly Copied from the source material, 17% were copied with a few Word Changes, 12% were copied with Grammar Changes and 3% appeared to be an ultimately unsuccessful attempt at paraphrasing (Complex).

For the most part, students who Directly Copied text into one area of their papers made an attempt at paraphrasing in other areas. Only 3 of the 32 students with plagiarism issues in their papers had only instances of Direct Copying in their papers. Most had a mix of many instances of Direct Copying with a few instances of Word Change or Grammar Change. Overall, 65% of students with incorrectly appropriated text had at least one sentence that was Directly Copied in their proposals. Sixty-one

percent had at least one sentence with a Word Change, 65% had at least one sentence with Grammar Changed and 27% had at least one sentence using a combination of paraphrasing techniques (Complex).

Extent of plagiarism (i.e. how many sentences were plagiarised and the percentage of plagiarised sentences in the entire paper) was related to the type of plagiarism detected. Generally, students who Directly Copied from a source did so many times and across a large percentage of their overall paper (Table 4). For example, the ten students with the highest percentage of plagiarism (over 20% of total paper content) also have the highest percentage of sentences that were Directly Copied from source material. All of the students that plagiarised more than 8% of their papers Directly Copied at least once. In contrast, of the 12 students that plagiarised 5% or less of their paper, only two had instances of Directly Copying the source material and only one had an instance of Word Change.

Table 4:

Number of plagiarised sentences by type ranked by the extent of plagiarism

Student #	Direct Copy	Word Change	Grammar Change	Complex	Total	% of total text plagiarised
1	31	2	1		34	65
2 (ESL)	26	2	2		30	37
3 (ESL)	23	5	2	1	31	32
4 (ESL)	10	5			15	30
5	12	3		2	17	26
6 (ESL)	22		2		2	24
7	11				11	24
8	17	12	2	2	33	23
9 (ESL)	29	3	5		37	21
10 (ESL)	21	12	2		35	21
11 (ESL)	5	1	3	1	10	19
12 (ESL)	8	4	4		16	17
13 (ESL)	6				6	14
14	1	5	1		7	13
15 (ESL)	23	2			25	12
16 (ESL)		3	7		10	9
17 (ESL)	3	1	1	1	6	9
18 (ESL)	3	2			5	8
19		1	1		2	7
20 (ESL)	2	1	1		4	6
21 (ESL)		1	2		3	4
22			1	1	2	4
23 (ESL)			2		2	4
24 (ESL)	3	1	1		5	3
25 (ESL)			1		1	2
26 (ESL)				2	2	2
27 (ESL)	1				1	2
28				1	1	1
29			3		3	1
30		1			1	1
31 (ESL)				1	1	1
32			1		1	1
Total	256	67	45	12	380	

Note: (ESL) designates students who self-identified or were identified by TOEFL scores as being non-native English speakers (English as a Second Language).

Types of source materials

Of the total pool of 380 plagiarised sentences detected, 30% of them were derived from popular technical sources such as patents, drug information sheets and lab protocols, 25% from primary journal articles, 24% from popular websites, 16% from secondary journal articles or books, and 5% from the abstracts of primary journal articles. Most of the students with plagiarised sentences in their papers used a combination of sources but those who mainly used popular technical sources plagiarised more sentences overall as compared to those who mainly used primary or secondary articles. Based on the total number of plagiarised sentences, it would appear that popular technical sources are the most commonly plagiarised source but when looking at the percentage of individual students who plagiarised from each source type, primary articles were most often plagiarised. Fifty-five percent of the students with plagiarism appropriated text from at least one primary journal article, 42% from a popular technical source, 33% from a popular website, 33% from a secondary source and 15% from the abstract of a primary journal article.

Citation status

Most instances of plagiarism were not cited at all or were cited incorrectly. Of the 380 plagiarised sentences, 80% had no citation, 13% were cited but incorrectly (10% cited the author who originally did the research while using a secondary article as a source and 3% cited the secondary author for work that someone else actually did) and only 7% were cited accurately. Although the vast majority of plagiarised sentences were not cited at all, 59% of the students made an attempt to cite at least some of their plagiarised sentences.

Location of plagiarised text

All of the research proposals submitted for this study contained Introduction and Method sections. Of the 32 proposals with plagiarism issues, only 9 writers wrote an Abstract while 23 included anticipated Results and 24 had Discussions. Plagiarism was most often detected in the Introduction section of the writing samples. Among the 32 students with at least one instance of plagiarism, all but one (97%) plagiarised in the Introduction of their paper. In contrast, 34% percent plagiarised in their Methods section. Of those students with plagiarism that produced abstracts, 22% had issues of plagiarism in this section. Eight percent of these students who included a Results section had issues of plagiarism in the results and 8% who included a Discussion section had issues of plagiarism in their discussion. When looking across the entire sample of plagiarised sentences, 75% of the 380 plagiarised sentences occurred in the Introduction of the research proposal and 20% occurred in the Methods section. When plagiarism was detected in the Results or Discussion sections it was usually just a sentence or two while entire paragraphs of the Introduction and Methods section were plagiarised in some cases. Plagiarism in the Introduction generally came from several sources while plagiarism in the Methods section usually came from a single research protocol found on the internet or in a journal article (upon comparison to the original source(s) it appears that 62% of the students with plagiarised sentences in their Methods sections took this information from only one source and 31% plagiarised two sources).

Interplay between type of plagiarism, type of source material and citation status

When comparing sentences that were Directly Copied, copied with Word Changes, and copied with Grammar Changes, several patterns emerge. Sentences that were Directly Copied were least likely to be cited correctly and most likely to come from popular websites or popular technical sources (Figure 1). Of the 256 examples of Directly Copied sentences, 84% of them were not cited anywhere in the paper, 13% were cited incorrectly and only 3% were cited correctly. Thirty-three percent of these sentences came from popular technical sources, 28% came from popular websites, 18% came from a primary journal article, 15% came from a secondary source and 6% came from the abstract of a primary source.

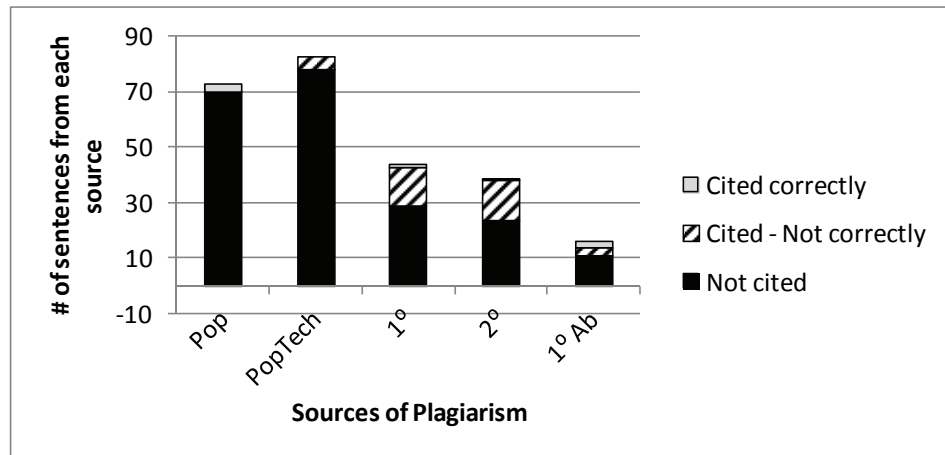


Figure 1: Sources used for sentences that were Directly Copied from the text (n=256 sentences)

The majority of sentences that were copied with Word Changes were not cited, but the percentage of correctly cited sentences was quadrupled as compared to cases of Direct Copying (Figure 2). Of the 67 sentences with Word Changes, 77% had no citation at all, 11% were cited incorrectly and 12% were cited correctly. Furthermore, a higher percentage of these sentences were taken from primary sources (35%) making that the most common source used for this type of plagiarism. Popular technical sources were also commonly used (29%) but the percentage of popular sources used was halved (14%) as compared to cases of Direct Copy.

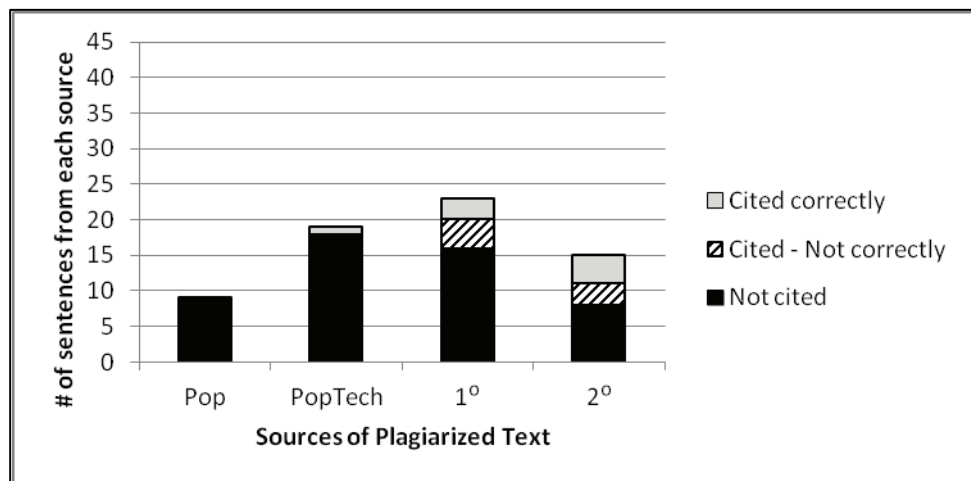


Figure 2: Sources used for sentences copied with Word Changes (n=67 sentences)

The majority of sentences that were copied with Grammar Changes were not cited but the percentage of correctly cited sentences increased greatly as compared to Direct Copy and Word Change (Figure 3). Of the 45 sentences with examples of this type of grammatical change, 71% were not cited anywhere in the paper, 9% were cited incorrectly and 20% were cited correctly. An even greater percentage of these sentences were taken from primary journal articles as compared to cases of Direct Copy or Word Change. Forty-seven percent came from primary sources, 24% from popular technical sources, 16% from popular websites, 11% from secondary sources and 2% from the abstract of a primary source.

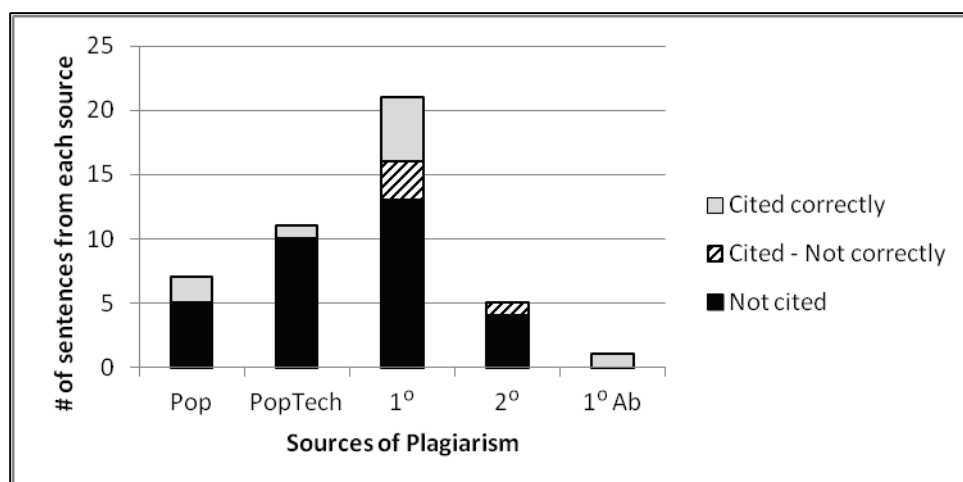


Figure 3: Sources used for sentences that were copied with Grammar Changes (n=45 sentences)

Only 12 sentences were coded as Complex and while a few were insufficiently paraphrased, most were coded as plagiarised because of a lack of citation. Patterns for Complex cases of plagiarism were very similar to that for Grammar Change. The majority (67%) of these sentences were not cited, 17% were cited incorrectly and 17% were cited correctly. The majority of these sentences were taken from primary sources (42%) while 25% were taken from popular websites, 17% from secondary sources and only 8% from popular technical sources.

When looking at the total sample of plagiarised sentences across all individuals it may at first appear that types of plagiarism, sources used and ability to correctly cite sources are all correlated. When looking within individuals, however, the patterns are less obvious in that nearly all the students committed more than one type of plagiarism and used a combination of source types (Table 4).

Patterns of plagiarism within individual students

Among the ten students with the greatest extent of plagiarism in their papers (those with 20% or more of the sentences in their paper coded as plagiarism), most of their problematic sentences were Directly Copied from a source and not cited, but many of these students did make an attempt to paraphrase and cited their sources in some instances (Figure 4). Those with a lesser extent of plagiarism in their paper (a smaller percentage of the sentences in their paper were coded as plagiarised) used a wider range of paraphrasing strategies and were more likely to try to cite their sources. In contrast, of the 12 students with less than 5% of the sentences in their paper plagiarised, only two had sentences that were Directly Copied from a source and the majority (67%) of them only had instances of Grammar Change or Complex paraphrasing. These students were also better at citing their sources. Thirty-three percent of them cited every source accurately and an additional 25% had a combination of accurately cited sources and incorrectly cited sources. Thirty-three percent of this group did not cite any of their inadequately paraphrased sentences however.

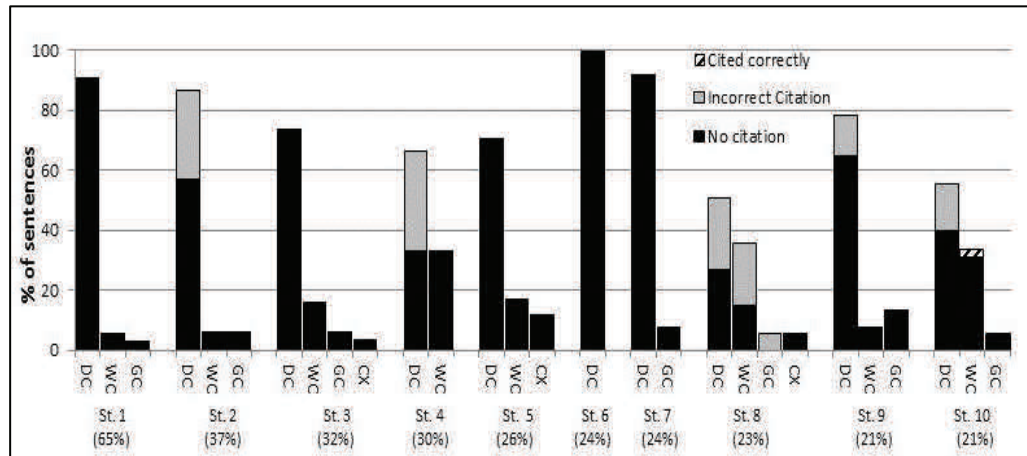


Figure 4: Type of plagiarism and citation errors by students that plagiarised more than 20% of their papers.

Of the ten students with the greatest percentage of plagiarism, some almost exclusively appropriated text from popular information found on the internet or popular technical sources while others used mostly primary and secondary sources (Figure 5). Most used a combination of sources. This pattern was also seen among the students with plagiarism issues in less than 5% of their papers, with 33% entirely appropriating text from primary sources while 42% relied heavily on popular websites. Regardless of whether a source was Directly Copied or attempts were made to paraphrase, sentences taken from primary or secondary sources were more likely to be cited than those taken from popular or technical web sources.

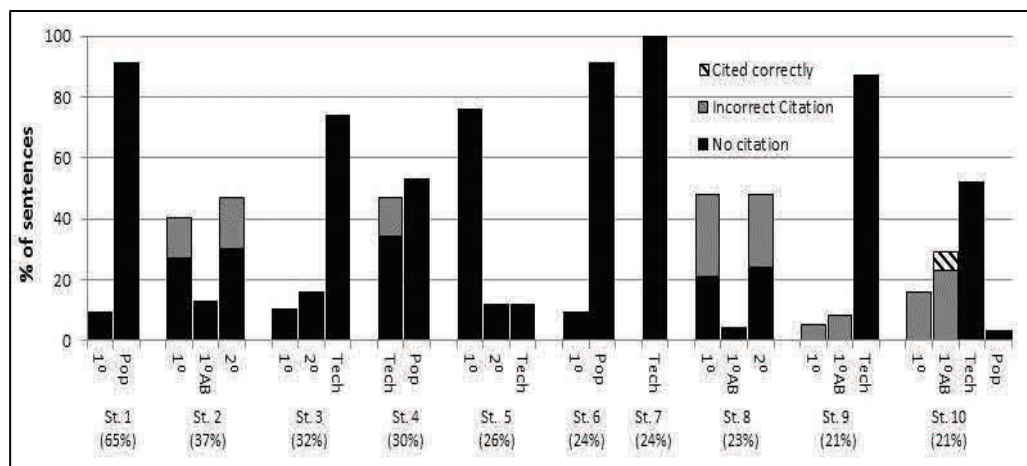


Figure 5: Sources used by the students that plagiarised more than 20% of their papers

Impact of being a non-native English speaker

Of the 115 students who submitted a research proposal, 39% reported that English was not their first language. Of these ESL students, almost half (47%) had at least one plagiarised sentence in their paper (compared to 16% of non-ESL students). Overall, a disproportionate number of the 32 papers containing plagiarism (65%) were written by ESL students.

Discussion

Most studies of plagiarism use self-report methodologies rather than empirical assessment, and data at the graduate level are particularly sparse; thus, our paper represents a significant contribution in this area. Although different methods were used to evaluate the extent of plagiarism, the overall rate of plagiarism in our study (28%) is similar to the rates reported for undergraduates (26% (Ellery, 2008) and 26–31% (Walker, 2010)) and master's theses (27% (McCullough & Holmburg, 2005)). Rates found in other graduate populations vary greatly. Five percent of medical residency applicants had at least one instance of plagiarism (Segal et al., 2010). In contrast, 76% of non-native English speaking graduate students had at least one passage in a writing sample (half of which were completed PhD theses) where over 70% of the words were taken from source material (Pecorari, 2003). Variation in rates of plagiarism reported may largely be due to the methodologies used to detect plagiarism as well as the nature of the writing task. Because an application for medical residency and a dissertation represent vastly different writing tasks (residency applications contain personal statements, while theses and dissertations are 'fact heavy'), it would seem that students are more likely to plagiarise when writing about highly factual content.

When they did provide a citation, the students in our sample rarely cited popular sources; more attempts were made to cite primary and secondary sources. When Ellery (2008) interviewed undergraduate plagiarisers she found many students held the misconception that online materials did not need to be cited. In contrast, her students expressed a greater awareness of the need for citation when referencing print materials, perceiving print articles as being more 'special' (Ellery, 2008). Our citation findings suggest that despite being able to access these primary and secondary sources online, at least some students are distinguishing between less 'special' general webpages and more 'special' primary sources.

Our sample contained a higher incidence of plagiarism in students for whom English is not a native language. Segal and colleagues (2010) found that plagiarism was more likely in medical residency applications of non-US citizens and residents who attended medical schools outside of the US. While citizenship and attending school outside the US do not necessarily mean that English was not a native language for those students, the percentage of non-native English speakers was likely higher in those two groups. Deckert (1993) surveyed Chinese undergraduates and found them to be unfamiliar with Western notions of plagiarism. He further interprets the Chinese cultural priority of conforming to authority as being a primary source of tolerance for plagiarism. In our study we equate primary literature as more authoritative than popular materials. We do not know the cultural backgrounds of our students, but as the type of source materials plagiarised are similar between ESL and native English speakers, cultural misunderstanding seems to be a less persuasive explanation. Maxwell, Curtis and Vardanega (2008) found no difference between students' abilities to recognise cases of plagiarism, and Pecorari (2003) argued both analytically and empirically against cultural influence as a cause of plagiarism.

Implications for practice

Beyond these more simplistic comparisons, the patterns among different text re-use characteristics provide empirical support for patchwriting as a developmental strategy (Howard, 1995). While no clear development trajectory was manifest, comparison of the students with the greatest percentage of text that was plagiarised with those with the smallest percentage does reveal a pattern in terms of how and from where they plagiarised. Specifically, students with the highest rates of plagiarism were far more likely to Directly Copy or Change Words than to attempt to paraphrase. They also drew mostly from popular or technical popular sites compared to students who made a greater attempt at paraphrasing and drew primarily from primary and secondary sources. These findings support a patchwriting framework wherein novice writers

must borrow large chunks of text and tend to seek out less sophisticated types of texts. More advanced writers, as they gain familiarity and comfort with the genre, may seek out more scholarly literature, begin to avoid popular sources and are able to summarise others' words more adeptly.

Additionally, the much higher incidence of plagiarism in the Introduction section of students' papers compared to the other sections also suggests that discomfort with content may contribute to plagiarism, even among experienced writers. Roig (2001) found that university professors were more likely to appropriate text from harder to understand sources. The Introduction section can be viewed as the most genre specific and perhaps intellectually challenging section for novice writers because it requires not only digestion and expression of ones' own work, but analysis and synthesis of other's work as well. Howard et al. (2010) examined summarising, paraphrasing, patchwriting, and copying in undergraduate writing. In the 18 writing samples they reviewed, they found no examples of summarisation, raising questions about how well students understood the sources from which they are writing (Howard et al., 2010). That some students borrowed heavily from the abstracts of primary works may indicate that they are not yet comfortable summarising the findings of primary works within their fields.

Lastly, as Pecorari discusses extensively (2003, 2006), non-native English speakers do not appear more likely to plagiarise than native speakers for any reason other than acquiring skill and comfort in academic writing is additionally challenging for them. Therefore, non-native English speakers should be offered additional support and unless there is specific evidence of intentional deception, all students should receive formative (non-punitive) feedback when plagiarism is detected.

Overall, our results provide further support for student use of patchwriting strategies at the graduate level, as well as the need for formative and explicit instruction regarding appropriate source use. The patterns of textual plagiarism characteristics can be used to guide and inform this instruction. Future research from this data source will allow us to examine in more detail the role of patchwriting in the developmental trajectory of student writing skill development, both by exploring the interactions between these characteristics of student plagiarism and student expertise, and by examining these patterns across multiple iterations of student writing.

The work presented in the paper was supported by a grant from the National Science Foundation (NSF-0723686) under the REESE program.

References

- Abasi, A., & Graves, B. (2008). Academic literacy and plagiarism: Conversations with international graduate students and disciplinary professors. *Journal of English for Academic Purposes*, 7, 221–223.
- Blackboard Inc. (2010). *Blackboard Learn™ Version 9.1 SafeAssign Instructor Guide*. Washington, DC: Blackboard Inc.
- Breen, L., & Maassen, M. (2005). Reducing the incidence of plagiarism in an undergraduate course: The role of education. *Issues in Educational Research*, 15(1), 1–16.
- Cole, A. F. (2007). Plagiarism in graduate medical education. *Family Medicine*, 39(6), 436–438.
- Couzin-Frankel, J., & Grom, J. (2009). Plagiarism sleuths. *Science*, 324, 1004–1007.
- Crocker, J., & Shaw, P. (2002). Research student and supervisor evaluation of intertextuality practices. *Hermes, Journal of Linguistics*, 28, 39–58.
- Deckert, G. D. (1993). Perspectives on plagiarism from ESL students in Hong Kong. *Journal of Second Language Writing*, 2(2), 131–48.

- DeVoss, D., & Rosati, A. C. (2002). "It wasn't me, was it?" Plagiarism and the web. *Computers and Composition*, 19, 191–203.
- Eret, E., & Gokmenoglu, T. (2010). Plagiarism in higher education: A case study with Prospective academicians. *Procedia Social and Behavioral Sciences*, 2, 3303–3307.
- Ellery, K. (2008). An investigation into electronic-source plagiarism in a first-year essay assignment. *Assessment & Evaluation in Higher Education*, 33(6), 607–617.
- Flowerdew, J., & Li, Y. (2007). Language re-use among Chinese apprentice scientists writing for publication. *Applied Linguistics*, 28(3), 440–465.
- Franklyn-Stokes, A., & Newstead, S. E. (1995). Undergraduate cheating: Who does what and why? *Studies in Higher Education*, 20(2), 159–172.
- Gilmore, J., Strickland, D., Timmerman, B., Maher, M., & Feldon, D. (2010). Weeds in the flower garden: An exploration of plagiarism in graduate students' research proposals and its connection to enculturation, ESL, and contextual factors. *International Journal of Educational Integrity*, 6(1), 13–28.
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12, 436–445.
- Harwood, N. (2009). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics*, 41, 497–518.
- Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57(7), 788–806.
- Howard, R. M. (1999). *Standing in the shadow of giants: Plagiarists, authors, collaborators*. Stamford, CT: Ablex Publishing Corporation.
- Howard, R. M., Serviss, T., & Rodrigue, T. K. (2010). Writing from sources, writing from sentences. *Writing & Pedagogy*, 2(2), 177–192.
- Jensen, L., Arnett, J., Feldman, S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology*, 27(2), 209–228.
- Johnson, R. L., Penny, J., Gordon, B., Shumate, S. R., & Fisher, S. P. (2005). Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores? *Language Assessment Quarterly: An International Journal*, 2(2), 117–146.
- Love, P. G., & Simmons, J. M. (1997, November 6-9). The meaning and mediated nature of cheating and plagiarism among graduate students in a college of education. Paper presented at the *Association for the Study of Higher Education Meeting*, Albuquerque, New Mexico.
- Mansourizadeh, K., & Ahmad, U. K. (2011). Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes*, 10, 152–161.
- Maxwell, A., Curtis, G. L., & Vardanega, L. (2008). Does culture influence understanding and perceived seriousness of plagiarism? *International Journal for Educational Integrity*, 4(2), 25–40.
- McCabe, D. L. (2005). Cheating among college and university students: A North American perspective. *The International Journal for Educational Integrity*, 1(1).
- McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior*, 11(3), 219–232.
- McCullough, M., & Holmberg, M. (2005). Using the Google search engine to detect word-for-word plagiarism in Master's theses: A preliminary study. *College Student Journal*, 39, 435–441.
- Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*, 12, 317–345.
- Pecorari, D. (2006). Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes*, 25, 4–29.
- Pecorari, D. (2008). *Academic writing and plagiarism: A linguistic analysis*. New York: Continuum.

- Petrić, B. (2007). Rhetorical functions of citations in high- and low-rated master's theses. *Journal of English for Academic Purposes*, 6, 238–253.
- Pickard, J. (2006). Staff and student attitudes to plagiarism at University College Northampton. *Assessment and Evaluation in Higher Education*, 31(2), 215–232.
- Roig, M. (1999). When college students' attempts at paraphrasing become instances of potential plagiarism. *Psychological Reports*, 84, 973–982.
- Roig, M. (2001). Plagiarism and paraphrasing criteria of college and university professors. *Ethics & Behavior*, 11(3), 307–323.
- Roig, M. (2009). Plagiarism: Consider the context. *Science*, 325, 813–814.
- Salager-Meyer, F. (1999). Referential behavior in scientific writing: A diachronic study (1810±1995). *English for Specific Purposes*, 18(3), 279–305.
- Segal, S., Gelfand, B. J., Hurwitz, S., Berkowitz, L., Ashley, S. W., Nadel, E. S., & Katz, J. T. (2010). Plagiarism in residency application essays. *Annals of Internal Medicine*, 153(2), 112–121.
- Selwyn, N. (2008). "Not necessarily a bad thing ...": A study of online plagiarism amongst undergraduate students. *Assessment and Evaluation in Higher Education*, 33(5), 465–479.
- Sheard, J., Dick, M., Markham, S., McDonald, I., & Walsh, M. (2002). Cheating and plagiarism: Perceptions and practices of first year IT students. *Proceedings of the 7th annual conference on Innovation and technology in computer science education*, 183–187.
- Simon, C. A., Carr, J. R., McCullough, S. M., Morgan, S. J., Oleson, T., & Ressel, M. (2004). Gender, student perceptions, institutional commitments and academic honesty: Who reports in academic dishonesty cases? *Assessment and Evaluation in Higher Education*, 29(1), 75–90.
- Sutherland-Smith, W. (2005). Pandora's box: Academic perceptions of student plagiarism in writing. *Journal of English for Academic Purposes*, 4, 83–95.
- Walker, J. (2010). Measuring plagiarism: Researching what students do, not what they say they do. *Studies in Higher Education*, 35(1), 41–59.
- Yeo, S. (2007). First-year university science and engineering students' understanding of plagiarism. *Higher Education Research & Development*, 26(2), 199–216.

Author Information

Michelle Vieyra PhD is an assistant professor of biology at the University of South Carolina Aiken. She teaches several writing intensive biology courses and has given workshops to faculty on her campus regarding student writing across the curriculum and the use of plagiarism detection software. Her research interests include the development of paraphrasing and citation skills from secondary through graduate school.

Denise Strickland is the Managing Director of the Center on Education Policy and Workforce Competitiveness in the Curry School of Education at the University of Virginia. A former high school teacher and biology researcher, her work in the area of science writing skill development has led to her current research interests focusing on enculturation into genres and how individuals learn to communicate effectively across discourses.

Briana Timmerman PhD is currently the Director of the Office of Teacher Effectiveness at the South Carolina Department of Education and Research Associate Professor in the Department of Biological Sciences at the University of South Carolina. Her research interests focus on how students of any age acquire inquiry skills such as scientific reasoning and scholarly writing.