

ORIGINAL ARTICLE

Open Access



CrossMark

# Using Internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism?

Ann M. Rogerson\* and Grace McCarthy

\* Correspondence:  
annr@uow.edu.au  
Faculty of Business, University of  
Wollongong, Building 40,  
Northfields Avenue, Wollongong,  
NSW 2522, Australia

## Abstract

A casual comment by a student alerted the authors to the existence and prevalence of Internet-based paraphrasing tools. A subsequent quick Google search highlighted the broad range and availability of online paraphrasing tools which offer free 'services' to paraphrase large sections of text ranging from sentences, paragraphs, whole articles, book chapters or previously written assignments. The ease of access to online paraphrasing tools provides the potential for students to submit work they have not directly written themselves, or in the case of academics and other authors, to rewrite previously published materials to sidestep self-plagiarism. Students placing trust in online paraphrasing tools as an easy way of complying with the requirement for originality in submissions are at risk in terms of the quality of the output generated and possibly of not achieving the learning outcomes as they may not fully understand the information they have compiled. There are further risks relating to the legitimacy of the outputs in terms of academic integrity and plagiarism. The purpose of this paper is to highlight the existence, development, use and detection of use of Internet based paraphrasing tools. To demonstrate the dangers in using paraphrasing tools an experiment was conducted using some easily accessible Internet-based paraphrasing tools to process part of an existing publication. Two sites are compared to demonstrate the types of differences that exist in the quality of the output from certain paraphrasing algorithms, and the present poor performance of online originality checking services such as Turnitin® to identify and link material processed via machine based paraphrasing tools. The implications for student skills in paraphrasing, academic integrity and the clues to assist staff in identifying the use of online paraphrasing tools are discussed.

**Keywords:** Paraphrasing, Internet tools, Plagiarism, Machine translation, Patchwriting, Academic integrity, Paraphrasing tools, Turnitin

## Introduction

A casual question from a student regarding another student's contribution to a group work assignment inadvertently led to an explanation of some unusual text submitted for assessment in a previous session. The student queried whether the use of a paraphrasing tool was acceptable in the preparation of a written submission for assessment. Discussing the matter further, the student revealed that they had queried the writing provided by one member of the group as their contribution to the report "did not make sense". When asked, the group member stated that they had taken material from

a journal article and used a fee free Internet paraphrasing tool “so that the words were not the same as the original to avoid plagiarism”. After the clarification, the group did not accept the submission from their team member and instead worked with them to develop an original submission. The group were thanked for their approach to the situation; however this revelation provided a potential explanation for some analogous submissions for previous subjects.

One particular submission from a previous subject instance had phrasing that included “constructive employee execution” and “worker execution audits” for an assessment topic on employee performance reviews. The student was interviewed at the time about why they had submitted work relating the words execution and employees and no satisfactory or plausible explanation was provided. With a new awareness of paraphrasing tools, a Google search revealed in excess of 500,000 hits and a simple statement was entered into one tool to test this connection. Testing the phrase ‘employee performance reviews’ via the top search response revealed an explanation for the unusual student submission as the paraphrase was returned as ‘representative execution surveys’. Choosing to use output generated by these tools begs the question – is it original work, patchwriting or facilitated plagiarism?

Having had our attention drawn to the existence and use of paraphrasing tools it was decided to investigate the phenomenon. What became apparent was that the ease of access to and use of such tools was greater than first thought. Consequently it is important to bring the use and operation of paraphrasing tools to a wider audience to encourage discussion about developing individual writing skills and improve the detection of these emerging practices, thereby raising awareness for students, teachers and institutions.

### **Paraphrasing and patchwriting**

Academic writing is largely reliant on the skill of paraphrasing to demonstrate that the author can capture the essence of what they have read, they understand what they have read and can use the appropriately acknowledged evidence in support of their responses (Fillenbaum, 1970; Keck, 2006, 2014; Shi, 2012). In higher education a student’s attempts at paraphrasing can provide “insight into how well students read as well as write” (Hirvela & Du, 2013, p.88). While there appears to be an underlying assumption that students and researchers understand and accept that there is a standard convention about how to paraphrase and appropriately use and acknowledge source texts (Shi, 2012), there can be inconsistencies between underlying assumptions in how paraphrases are identified, described and assessed (Keck, 2006). Poorer forms of paraphrasing tend to use a simplistic approach where some words are simply replaced with synonyms found through functionality available in word processing software or online dictionaries. This is a form of superficial paraphrasing or ‘close paraphrasing’ (Keck, 2010) or ‘patchwriting’ (Howard, 1995). The question as to “the exact degree to which text must be modified to be classified as correctly paraphrased” (Roig, 2001, p.309) is somewhat vague, although Keck (2006) outlined a *Taxonomy of Paraphrase Types* where paraphrases are classified in four categories ranging from near copy to substantial revision based on the number of unique links or strings of words.

Research in this area appears to concentrate more specifically on second language (L2) students rather than students per se (For a review see Cumming et al. 2016)

although many native English writers may also lack the language skills to disseminate academic discourse in their own voice (Bailey & Challen, 2015). Paraphrasing is a skill that transcends the written form as it is actually a communication strategy required for all language groups in interpersonal or intergroup interactions and includes oral (Rabab'ah, 2016) and visual forms (Chen et al. 2015a). Paraphrasing allows the same idea to be expressed in different ways as appropriate for the intended audience. It can also be used for persuasion (Suchan, 2014), explanations (Patil & Karekatti, 2015) and support (Bodie et al. 2016). In coaching, paraphrasing is used to ensure that the coach has correctly understood what the coachee is saying, thus allowing the coachee to further clarify their meaning (McCarthy, 2014).

### **Online writing tools**

The prevalence and easy access to digital technologies and Internet-based sources have shifted "the way knowledge is constructed, shared and evaluated" (Evering & Moorman, 2012, p.36). However the quality, efficacy, validity and reliability of some Internet-based material is questionable from an educational standpoint (Niño, 2009). Internet-based paraphrasing tools are text processing applications and associated with the same approaches used for machine translation (MT). While MT usually focusses on the translation of one language to another, the broader consideration of text processing can operate between or within language corpuses (Ambati et al. 2010).

Internet-based conversion and translation tools are easily accessible, and a number of versions are available to all without cost (Somers, 2012). Developments in the treatment of translating natural language as a machine learning problem (known as statistical machine translation - SMT) are leading to continual improvements in this field although the linguistic accuracy varies based on the way each machine 'learns' (Lopez, 2008). The free tools available via the Internet lack constant updates and improvements as the code is controlled by webmasters and not by experts in MT (Carter & Inkpen, 2012). This means advances in methods and algorithms are not always available to individuals relying on free Internet based tools. Consequently there are issues with the quality of MT which may require a level of post-editing to correct the raw output so that it is fit for purpose (Inaba et al. 2007).

Post-editing of an online output may be problematic or difficult for an individual with a low level of proficiency in the language they are being taught or assessed in as grammatical inaccuracies and awkward phrasing cannot be easily identified and therefore corrected (Niño, 2009). Where a student is considered to lack the necessary linguistic skills, the errors or inaccuracies may be interpreted by assessors as a student having a poor understanding of academic writing conventions rather than recognising that a student may not have written the work themselves. Where an academic is working in an additional language, they may find the detection of the errors or inaccuracies more difficult to identify.

Nor is the issue of paraphrasing or article spinning tool use confined to students. Automated article spinners perform the same way as paraphrasing tools, where text is entered into one field with a 'spun' output provided on the same webpage. They were initially developed for re-writing web content to maximise exposure and links to particular sites, without being detected as a duplicate of original content (Madera et al. 2014). The underlying purpose appears to allow website owners to "make money from

the new, but not strictly original, article" (Lancaster & Clarke, 2009). These sites are freely available to students leading to a new label covering the use of these tools as 'essay spinning' (Lancaster & Clarke, 2009, p.26). However, these spinning tools are equally available to academics who may be enticed with the notion of repurposing already published content as a way of increasing research output.

Although the quality levels of MT output varies widely, careful editing and review can address the errors further disguising the original source material (Somers, 2012). Roig (2016) highlights that some forms of text recycling are normal in academic life such as converting conference presentations and theses to journal articles and the textual reuse between editions of books, as long as there is appropriate acknowledgement of the original source. However Roig also points out that authors should be concerned about reusing previous work as with technological advances it will not be long before all forms of academic written work can "be easily identified, retrieved, stored and processed in ways that are inconceivable at the present time" (Roig, 2016, p.665).

The fact remains that taking another author's work, processing it through an online paraphrasing tool then submitting that work as 'original' is not original work where it involves the use of source texts and materials without acknowledgement. The case of a student submitting work generated by an online tool without appropriate acknowledgement could be considered as a form of plagiarism, and the case of academics trying to reframe texts for alternate publications could be considered as a form of self-plagiarism. Both scenarios could be considered as 'facilitated plagiarism' where an individual actively seeks to use some form of easily accessible Internet-based source to prepare or supplement submission material for assessment by others (Granitz, 2007; Scanlon & Neumann, 2002; Stamatatos, 2011). Applying technology to identify where the paraphrasing tools have been used is difficult as detection moves beyond text summarisation and matching to comparison of meaning and evaluation of machine translation (Socher et al. 2011).

Furthermore, students using an online paraphrasing system fail to demonstrate their understanding of the assessment task and hence fail to provide evidence of achieving learning outcomes. If they do not acknowledge the source of the text which they have put through the paraphrasing tool, they are also guilty of academic misconduct. On both counts, they would not merit a pass in the subject for which they submit such material.

### **Methodology**

In order to test the quality of output generated by some free Internet based paraphrasing tools and how the originality of the output is assessed by Turnitin®, the following experiment was conducted. A paragraph from an existing publication by this article's authors from a prior edition of the International Journal of Educations Integrity (IJEI) was selected to be the original source material (McCarthy & Rogerson, 2009, p.49). To assess how a paraphrasing tool processes an in-text citation, one in-text citation was included (Thatcher, 2008). A set of three bibliographic entries from the reference list of the same article were also selected to test how references are interpreted.

As students are more likely to use Google as the Internet search engine of choice and rely on results near the top of page (Spievak & Hayes-Bohanan, 2016), this approach was used to identify and select some online paraphrasing tools for testing. The selected

paragraph (including the in-text citation), and the selected references were entered into the first two hits on a Google search on www.google.com.au for '*paraphrasing tools*'. Consequently the sites used for the experiment were www.paraphrasing-tool.com (Tool 1) and www.goparaphrase.com (Tool 2).

The next step was to compare the outputs from the original journal article material to the outputs of Tool 1 and Tool 2. Exact matches to the original text were observed, tagged and highlighted in grey. Matches between the two paraphrasing outputs that did not match the original source were highlighted by placing the relevant text in a box. Contractions and unusual matches were highlighted by double underlining the text. For the first set of comparisons (paragraph with an in-text citation) the following summary characteristics were calculated: total word counts, total word matches and percentage of similarity to the original paragraph.

In order to identify how Turnitin® interpreted the paragraph and bibliographic outputs from the paraphrasing tools, the original source material and two paraphrasing outputs were uploaded to Turnitin® to check whether the journal publication could be identified. Turnitin® comprises a suite of online educative writing and evaluation tools where assessment tasks can be uploaded, checked and assessed ([www.turnitin.com](http://www.turnitin.com)). It can be accessed via the Internet or through an interface with an institutional learning management system (LMS). The originality checking area compares a submission against a range of previously published materials and a database of previously submitted assignments. The system generates an originality report where text that matches closely to a previously published or submitted source is highlighted by colour and number with links provided to publicly accessible materials. Matches to papers submitted at other institutions cannot be accessed without the express permission of the owning institution. As Baggaley and Spencer note (2005) Turnitin® originality reports require careful analysis, for the reports identify text "which may or may not have been correctly attributed" (Baggaley & Spencer, 2005, p. 56) and cannot be used as the sole determinant of whether or not a work is plagiarised or if source materials have been inappropriately used (Rogerson, 2014).

A separate Turnitin® assessment file was created for the experiment on an institutional academic integrity LMS site (Moodle) where a bank of dummy student profiles is available for testing purposes. Three dummy student accounts were used to load the individual '*outputs*' under two assignment parts. The uploads included one instance of the source material in order to generate comparative originality reports for both the paragraph outputs (loaded under part 1) and the reference list outputs (loaded under part 2). For both sets of outputs the overall Turnitin® similarity percentages and document matches were reviewed for comparison purposes.

## Results

The highlighted comparisons of the paragraph outputs are presented in Fig. 1 (comparing Tool 1) and Fig. 2 (comparing Tool 2). The summary characteristics for the paragraph outputs are presented in Table 1.

There are obvious differences in how the online paraphrasing tools have reengineered the original work based on the number of identifiable matches between the original and output texts. For example there are differences in how words such as plagiarism are expressed (Original source: *plagiarism*; Tool 1: *copyright infringement*; Tool 2:

Original source:	Output from Tool 1: <a href="http://www.Paraphrasing-tool.com">www.Paraphrasing-tool.com</a>
<p>The ease of access to information sources does not guarantee that students allocate time to review governance requirements. This same ease of access is sometimes seen as the underlying cause for increases in events that have the potential to be classified as, or may in fact be, plagiarism. For many overseas students, the actual concept of plagiarism is not readily understood, as copyright laws and the requirement to attribute ideas to originating authors is not promoted, policed, or in many cases does not exist in their home countries or institutions. Other nationalities can have a different perspective regarding the sharing of ideas, or have different words to describe plagiarism that may or may not have the same implication of unacceptable behaviour as it does in the English language. Thatcher (2008) notes that rather than being unacceptable behaviour, the Chinese regard copying as a way of learning from and paying respect to past masters.</p> <p>(McCarthy &amp; Rogerson, 2009, p.49)</p>	<p>The straightforward entry to data sources does not ensure that undergraduates apportion time to audit administration necessities. This same straightforward entry is in some cases seen as the basic reason for increments in occasions that can possibly be named, or may in actuality be, copyright infringement. For some abroad undergraduates, the real idea of literary theft is not promptly comprehended, as copyright laws and the prerequisite to ascribe thoughts to starting creators is not advanced, policed, or much of the time does not exist in their nations of origin or organizations. Different nationalities can have an alternate point of view with respect to the sharing of thoughts, or have diverse words to portray unoriginality that might possibly have the same ramifications of inadmissible conduct as it does in the English dialect. Thatcher (2008) takes note of that instead of being unsuitable conduct, the Chinese see replicating as a method for gaining from and paying appreciation to past experts.</p>

**Fig. 1** Comparison with output from [www.paraphrasing-tool.com](http://www.paraphrasing-tool.com). Original source materials from McCarthy and Rogerson (2009, p.49) and citing Thatcher (2008)

*counterfeit*). Both tools have used additional words (Tool 1: additional five words; Tool 2: additional 20 words). The output from Tool 1 has used 77 words or 50% of the words in the original paragraph but these were predominately coordinating conjunctions. Tool 1 has followed the correct use of capitalisations in all words and sentences, however Tool 2 has not capitalised words such as English, and Chinese, but did capitalise seven random words mid-sentence (*Audit*, *Numerous*, *Concerning*, *Likewise*, *Taking*, and *What's*). In addition Tool 2 used contractions (*doesn't*) and the words '*can have*' in the original have been reprocessed to '*camwood*'.

The highlighted comparisons of the reference section outputs are presented in Fig. 3 (comparing the original source with Tool 1 and Tool 2). The summary characteristics of the Turnitin® results for the reference section outputs are presented in Table 2.

The Turnitin® results for both the paragraph and reference list uploads identified the original source as 100% match to the online location of the journal supporting Turnitin's® claim in relation to identifying legitimate academic resources. What is of concern is Turnitin's® apparent inability to identify the similarities evident by a manual comparison of the source and outputs. Figures 1 and 2 demonstrate the similarities between the original source materials and the output of the tools yet the similarity percentages noted in Table 2 indicate that the re-engineered paragraphs are not detected. One of the current limitations of Turnitin® is that it can detect some but not all cases of synonym replacement (Menai, 2012). Despite the patterned nature

Original source:	Output from Tool 2: <a href="http://www.Go-paraphrase.com">www.Go-paraphrase.com</a>
<p>The ease of access to information sources <u>does not guarantee</u> that students allocate time to review governance requirements. This same ease of access is sometimes seen as the underlying cause for increases in events that have the potential to be classified as, or may in fact be, plagiarism. For many overseas students, the actual concept of plagiarism is not readily understood, as copyright laws and the requirement to attribute ideas to originating authors is not promoted, policed, or in many cases <u>does not exist</u> in their home countries or institutions. Other nationalities can have a different perspective regarding the sharing of ideas, or have different words to describe plagiarism that may or may not have the same implication of unacceptable behaviour as it does in the English language. Thatcher (2008) notes that rather than being unacceptable behaviour, the Chinese regard copying as a way of learning from and paying respect to past masters.</p> <p>(McCarthy &amp; Rogerson, 2009, p.49)</p>	<p>Those simplicity about right with data wellsprings <u>doesn't</u> surely that people dispense the long run should Audit governance necessities. This same simplicity for right will be here and there seen. Similarly as the underlying make for expands over occasions that have the possibility will a chance to be arranged as, alternately might indeed be, counterfeiting. To Numerous abroad students, those real idea of counterfeiting is not <u>promptly</u> understood, as copyright laws and the <u>prerequisite</u> to quality plans with beginning <u>creators</u> will be not promoted, policed, or done huge numbers situations <u>doesn't</u> exist in their home nations alternately establishments. Different nationalities <u>camwood</u> need an alternate point of view in regards to the imparting from claiming ideas, alternately have distinctive expressions on describe counterfeiting that <u>might</u> alternately might not bring the same suggestion about unsuitable self-destructive considerations and conduct Concerning illustration it can in the <u>english</u> dialect. Thatcher (2008) notes that as opposed being unsuitability behaviour, the <u>chinese</u> view duplicating Likewise an approach about Taking in starting with What's more paying admiration to previous aces.</p>

**Fig. 2** Comparison with [www.goparaphrase.com](http://www.goparaphrase.com). Original source materials from McCarthy and Rogerson (2009, p.49) and citing Thatcher (2008)

of the text matching identified through a visual examination of the output, the machine-based originality similarity checking software continues to have limitations in identifying materials that appear to be plagiarised through the use of an online paraphrasing tool or language translation application.

Turnitin® was more successful in matching up bibliographic data to the original source. This was likely due to the fact that the paraphrasing tools did not alter (or barely altered) long strings of numbers, letters and website URLs. The higher Turnitin® match to the output from Tool 1 (72% similarity) was due to the retention of most of the journal name (*International* replaced with *Global*) however the author name '*Crisp*'

**Table 1** Comparison of paragraph elements

Source	Total words	Words the same as original paragraph	Percentage of word similarity to original paragraph
Original paper	153		
Tool 1	158	77	50%
Tool 2	173	46	30%

Original references
Carroll, J., & Appleton, J. (2005). Towards consistent penalty decisions for breaches of academic regulations in one UK university [Electronic Version]. <i>International Journal for Educational Integrity</i> , 1, 1-11, from <a href="http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/15/5">http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/15/5</a>
Crisp, G. T. (2007). Staff attitudes to dealing with plagiarism issues: Perspectives from one Australian university [Electronic Version]. <i>International Journal for Educational Integrity</i> , 3, 3-15, from <a href="http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/130/109">http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/130/109</a>
Dahl, S. (2007). Turnitin (R): The student perspective on using plagiarism detection software. <i>Active Learning in Higher Education</i> , 8(2), 173-191.
<b>Tool 1: Paraphrasing-tool.com</b>
1 Carroll, J., and Appleton, J. (2005). Towards reliable punishment choices for ruptures of scholarly controls in one UK college [Electronic Version]. <i>Global Journal for Educational Integrity</i> , 1, 1-11, from <a href="http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/15/5">http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/15/5</a>
Fresh, G. T. (2007). Staff mentalities to managing literary theft issues: Perspectives from one Australian college [Electronic Version]. <i>Global Journal for Educational Integrity</i> , 3, 3-15, from <a href="http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/130/109">http://www.ojs.unisa.edu.au/index.php/IJEI/article/viewFile/130/109</a>
Dahl, S. (2007). Turnitin 1: The understudy point of view on utilizing literary theft discovery programming. <i>Dynamic Learning in Higher Education</i> , 8(2), 173-191.
<b>Tool 2: Go paraphrase.com</b>
Carroll, J. , & Appleton, J. (2005). To 1 ds steady punishment choices to breaches about academic regulations Previously, you quit offering on that one uk college [Electronic Version]. Global diary for instructive Integrity, 1, 1-11, from <a href="http://www.Ojs.Unisa.Edu.Au/index.Php/IJEI/article/viewFile/15/5">http://www.Ojs.Unisa.Edu.Au/index.Php/IJEI/article/viewFile/15/5</a> .
Crisp, g. T. (2007). Staff attitudes will managing counterfeit 1 issues: Perspectives from particular case australian college [Electronic Version]. Worldwide diary to instructive Integrity, 3, 3-15, from <a href="http://www.Ojs.Unisa.Edu.Au/index.Php/IJEI/article/viewFile/130/109">http://www.Ojs.Unisa.Edu.Au/index.Php/IJEI/article/viewFile/130/109</a> .
2 Dahl, encountered with urban decay because of deindustrialization, engineering concocted, government lodg. (2007). Turnitin (R): the person viewpoint ahead utilizing counterfeiting identification programming. Animated Taking in higher Education, 8(2), 173-191.

**Fig. 3** Comparison with three reference list entries. Original source materials from McCarthy and Rogerson (2009, p.56) and citing Carroll and Appleton (2005), Crisp (2007), and Dahl (2007)

was altered to '*Fresh*'. The output from Tool 2 retained the authors' last names, but added in 11 additional words to replace author Dahl's first initial of 'S' which would have affected the calculation of similarity percentage. It is interesting that the change to lower case for authors' initials appeared to impact on Turnitin's® capacity to identify the authors in the first reference and missed the end of the journal details in the third reference, which also would have contributed to the lower similarity percentage. This led to Turnitin® overlooking 15 word matches and 13 other number and character matches in the Tool 2 submission that were identified as direct matches in the Tool 1 output.

**Table 2** Turnitin originality score comparisons

	Paragraph comparisons			Reference list comparisons		
	Turnitin® Paper ID	Similarity	Document Matches	Turnitin® Paper ID	Similarity	Document Matches
Original paper	715632730	100%	1	715635319	100%	1
Tool 1	715633115	0%	0	715635881	72%	1
Tool 2	715633512	0%	0	715636200	47%	2

A further examination of both sets of outputs from the paraphrasing tools identified that the tools appear to retain most words and formatting close to punctuation. For example both tools retained [, *policed*,], and the name and intext citation [Thatcher (2008)] in the paragraph comparison, and a string in the reference section comparison [*Integrity*, 3, 3–15, from <http://www.>]. Without knowing the algorithms for the paraphrasing tools or Turnitin®, patterns such as these can only be observed rather than analysed.

## Discussion

The outputs and comparisons presented in Figs. 1 and 2 appear more like patchwriting rather than paraphrasing. Li and Casanave (2012) argue that patchwriting is an indication that the student is a novice writer still learning how to write and understand the “complexities of appropriate textual borrowing” (Li & Casanave, 2012, p.177) although their study was confined to L2 students submitting assessment material in English. They further argue that deeming text as patchwriting does not attract the same negative connotations of plagiarism nor would it attract the same penalties. In our examples the patterns of text, language and phrasing can identify a student requiring learning support. This determination is likely due to the presence of poor expression, grammatical errors and areas of confused meaning which are sometimes referred to as a ‘word salad’. The term word salad is drawn from psychology but has been adopted in areas such as MT to classify unintelligible and random collections of words and phrases (Definition:word salad, 2016). Word salads are produced by MT “when translation engines fail to do a complete analysis of their input” (Callison-Burch & Flournoy, 2001, p.1).

While the output from Tool 1 is mainly intelligible, some of the results from Tool 2 could be classified as word salads, for example in the last line the following string of words was produced ‘*duplicating Likewise an approach about Taking in starting with What's more paying admiration to previous aces*’. If an unintelligible string of words was submitted as part of an assessment task it may be a reason to have a conversation with a student to understand how they are going about their writing, and to determine if paraphrasing tools or article spinners have contributed. Where a citation is provided, it may be a case of a student having a poor understanding of academic writing conventions. Where there is no citation or any reference to the original source the situation may warrant investigation under academic integrity institutional policies and procedures.

If the percentage calculations presented in Fig. 1 are compared with Kecks (2006) *Taxonomy of Paraphrase Types*, the outputs from the online tools would fall into the category of paraphrases with minimal revision when compared to the original text (Keck, 2014, p.9). The manual comparison of documents in this experiment indicates a level of patchwriting, however Turnitin® could not establish a relationship between the original source paragraph and the machine generated paraphrasing-tool outputs. It is more akin to some of the plagiarism behaviours described by Walker (1998, p.103) such as “illicit paraphrasing” where material is reused without any source acknowledgement or even “sham paraphrasing” where text is directly copied but includes a source acknowledgement. This is a cause for concern as the comparison with the online paraphrasing tool output was only possible as the original source was known. It is

not just a question of percentages but in the patterns clearly visible in Figs. 1, 2 and 3. Consequently, this set of experiments indicates a level of similarity that is concerning in two key areas, firstly where the original source is not acknowledged or identifiable, and secondly if this level of similarity were found in student work, it would suggest that the student may not have understood the material, or at least that he/she has not demonstrated their understanding.

Manual analysis and academic judgement are integral parts of the process of detection of plagiarised materials (Bretag & Mahmud, 2009b), and are heavily reliant on the level of experience an assessor has in identifying clues, markers and textual patterns (Rogerson & Bassanta, 2016). In this experiment the original source of the plagiarised materials would be difficult to identify, however the presence of clues and patterns may be sufficient to motivate a lecturer or tutor to initiate an initial conversation with a student to determine whether the work is actually the student's own (Somers et al. 2006).

A further investigation of the results from the Google search on 'paraphrasing tools' identified that many of the sites have multiple public faces—that is that there are additional URLs that direct users back to the same paraphrasing machine. The purpose behind the existence of the sites is not clear. The sites do carry Internet advertising so their existence and multiple faces may be related to a way to generate income. Alarmingly the sites examined in this study showed advertisements for higher education institutions which could be misinterpreted by users as tacit approval for the sites and their output. Other sites highlight that rudimentary paraphrasing tools are highly inaccurate but promote their paid services to correct the output—i.e. a process that could be interpreted as another form of contracted plagiarism (Clarke & Lancaster, 2013).

One of the questions that arises in assessing work as plagiarised is associated with intentionality—that is, did the person intend to deceive another about the originality of work (Lee, 2016). In the case of students "it is the inappropriate research and writing practices and the resulting misappropriate or misuse of information that leads students to breach academic integrity expectations" (Pfannenstiel, 2010, p.43). Pfannenstiel's use of the word 'expectations' is both interesting and enlightening as it is probable that differences in expectations is what is at the crux of the issue with online paraphrasing or article spinning tools. Expectations can be influenced by cultural and educational backgrounds, a lack of understanding or skills in paraphrasing and linguistic and language resources (Cumming et al., 2016; Sun, 2012). For example: a student may sincerely believe that as they have not submitted an exact copy of the original source, and that there is no evidence of match to the original source via online originality checking software that they have met the objective of submitting original work. Conversely, an academic may reasonably consider this to be direct plagiarism as the student copied the original work of someone else and reused it without any acknowledgement (Davis & Morley, 2015). This area of confusion was noted in Shi's (2012) study where a student stated that using a translation of an original text did not require acknowledgment of the original source as the translation was not directly the original source. (Shi, 2012, p.140).

While Turnitin® cannot currently connect the writing and the paraphrases in this experiment, it and other MT tools are in a constant state of evolution and their ability to identify poor quality machine translated text will continue to improve over time (Carter & Inkpen, 2012). In order to test the progress, Carter and Inkpen (2012) suggest that multiple tests of the same piece of text be conducted over a period of years to measure

both the quality of output and the ability to detect their use. The literature reviewed in this area focusses on the detection of phrases and sentences, with Socher et al. (2011) noting that once detection switches from phrases to full sentences a comparison of meaning is more difficult for a machine to learn.

This article does not attempt to outline all the work being undertaken in this area, instead it highlights that there is research being undertaken to develop and further enhance MT (encoding and decoding) and detection of MT use. This includes computers learning computational semantics and managing expanded vocabularies to move beyond recognition of specific tasks (Kiros et al., 2015). Turnitin®'s ability to match large sections of text outside of their own repository of previously submitted assessment tasks is very useful because the majority of academic materials that can be plagiarised are text based (Bretag & Mahmud, 2009a). Using text-matching as a basis for detection instead of semantic matching means that uses of online paraphrasing tools and article spinners continues to be difficult for technology to detect at this time. Therefore for the foreseeable future the onus of detection of unoriginal material remains with academics, lecturers and teachers (Rogerson, 2014).

Further confusion arises when institutions develop computer based paraphrasing tools as a way of developing English language writing skills for L2 students. Aware of the difficulties that L2 learners have with paraphrasing tasks, Chen, Huang, Chang and Liou developed a web and corpus based 'paraphrasing assistant system' designed to suggest paraphrases with corresponding Chinese translations (Chen et al. 2015b, p.23). Students familiar with using such a system in their home country may seek similar assistance if studying abroad. Without access to an approved technology they may seek to discover similar assistance tools on the Internet—where they can easily locate the paraphrasing tools identified in this experiment. These same students may also lack the judgement skills to discern the difference between the output from approved and poor quality online tools whether they are paraphrasing tools, article spinners or language translators.

### **Implications for practice: working with students**

One way of confronting or approaching this issue is to openly demonstrate to students the errors and inaccuracies that can result in using online tools (Niño, 2009). Communicating proactively about the issue provides students with a greater awareness of the problems that can result from using online paraphrasing sites as well as ensuring that students understand that they should not expect to graduate unless they can demonstrate they understand the course material. Their current and future employers have the right to expect that for example, a student graduating with a degree in marketing will be able to articulate their understanding of marketing concepts. Proactive approaches can also promote learning development and support services offered by the educational institution providing students with advice about paraphrasing and strategies for improving their writing skills and therefore avoiding problematic practices. This educates students about alternatives to using online machine text generation tools.

Some students have expressed concerns that other students will continue to take advantage of technology based aids even though they had been told not to use them and knowing that to do so could be classified as cheating (Burnett et al. 2016). Students who do not cheat but put in the effort themselves are usually outraged if fellow

students get away with cheating and may even bring cases they notice to the institutions' attention (Warnock, 2006). This was the case with the casual comment by the student who brought the online paraphrasing tools to our attention. The actions of our students working with their group member to develop their own work also demonstrates how honest students can be allies in upholding the academic standards of the institution (Bretag & Mahmud, 2016). If the benefits of learning and developing individual paraphrasing skills are linked to the broader benefits of effective interpersonal and intergroup communication, the open approach to confronting and discussing the issue may be more successful.

### **Implications for practice: working with staff**

The development of reading, summarising and paraphrasing skills are not the sole responsibility of learning developers. Educators need to embed academic skills in lectures and tutorials and provide feedback on student progress measured through effective assessment (Sambell et al. 2013). Clear assessment requirements and use of rubrics indicate the importance and differences to grades for the various levels of academic skills (Atkinson & Lim, 2013) providing students with a reason to develop their skills. Effective feedback assists students in identifying where they have achieved certain levels of academic skills and which skills require further development (Evans, 2013).

A further approach to tackling the issue is to re-design assessment tasks to include an oral component where the student has to present a summary of their argument and answer questions. This approach can ensure that the student understands and has achieved the learning outcomes, although it is no guarantee of the student's academic integrity in preparing for their presentation. Finally, academics can also be trained to look for linguistic markers indicating the possibility of the use of such online paraphrasing tools so that they can investigate cases appropriately. Such markers include sentences that do not make sense, odd use of capitalisations in the middle of sentences, unusual phrases and, in the case where students have reprocessed work from old textbooks, out of date and superseded reference material.

### **Conclusion and recommendations for further research**

This study has demonstrated that students can use online paraphrasing tools or article spinners in ways that avoid detection by originality checking software such as Turnitin®. Whether or not it is the student's intent to avoid plagiarism is not the issue examined here. Rather, the intent of this paper is to ensure that those involved in teaching and learning are aware of the practice, can detect its use and initiate meaningful conversations with students about the perils of using such tools. There is a fine line between use of paraphrasing tools and the use of tools to plagiarise, however it is only through open discussion that students will learn to appreciate the benefits of articulating their understanding in their own words with the appropriate acknowledgement of sources.

Paraphrasing is a skill that transcends an ability to interpret and restate an idea or concept in writing. It is an important skill that needs to be introduced and developed in terms of written, visual and oral forms. The capacity of students and academics to rephrase, frame and restate the ideas and intentions of original authors themselves with appropriate acknowledgements of sources is fundamental to the principles of academic integrity and personal development. The proliferation of fee-based and free

Internet-based tools designed to re-engineer text is a concern. Of greater concern is that tools contracted to identify original source materials cannot necessarily be used at this time to identify where writing has been repurposed. Regardless of the ease of access to online text regeneration tools and the work being done to try to electronically detect their use, individuals should be encouraged to improve their own paraphrasing expertise as an essential part of individual skill development in and beyond educational institutions.

Further work is needed to identify linguistic markers indicating use of online paraphrasing tools such as those identified in this study. Academics are already time poor and while they may be strongly in favour of upholding academic standards, they may also be reluctant to undertake time-consuming investigations into possible misconduct. They need encouragement to integrate the observation of textual patterns and markers into their grading and assessment practice. Research is also needed in exploring the most effective techniques or combination of educational, deterrent and punitive techniques and machine detection tools to combat the use of online paraphrasing tools and article spinners and other forms of academic malpractice. Such developments will assist in directing the focus of writing efforts back to where it should be – which is individuals writing and submitting their own work with appropriate acknowledgements.

#### **Acknowledgments**

The authors would like to thank the two anonymous reviewers for their constructive feedback on the original version of this manuscript.

#### **Authors' contributions**

AR 80%. GM 20%. Both authors read and approved the final manuscript.

#### **Competing interests**

The authors declare that they have no competing interests.

Received: 9 October 2016 Accepted: 10 December 2016

Published online: 26 January 2017

#### **References**

- Ambati V, Vogel S, and Carbonell JG (2010). Active learning and crowd-sourcing for machine translation. Paper presented at the Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Malta.
- Atkinson D, Lim SL (2013) Improving assessment processes in Higher Education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australas J Educ Technol* 29(5):651–666
- Baggaley J, Spencer B (2005) The mind of a plagiarist. *Learning, Media and Technology* 30(1):55–62. doi:10.1080/13581650500075587
- Bailey C, Challen R (2015) Student perceptions of the value of Turnitin text-matching software as a learning tool. *Practitioner Research in Higher Education* 9(1):38–51
- Bodie GD, Cannava KE, Vickery AJ (2016) Supportive communication and the adequate paraphrase. *Commun Res Rep* 33(2):166–172. doi:10.1080/08824096.2016.1154839
- Bretag T, and Mahmud S (2009a). A model for determining student plagiarism: Electronic detection and academic judgement. Paper presented at the 4APFEI Asia Pacific Conference on Education Integrity APFEI, Wollongong.
- Bretag T, Mahmud S (2009b) Self-plagiarism or appropriate textual re-use? *J Academic Ethics* 7(3):193–205. doi:10.1007/s10805-009-9092-1
- Bretag T, Mahmud S (2016) A conceptual framework for implementing exemplary academic integrity policy in Australian higher education. In: Bretag T (ed) *Handbook of Academic Integrity*. Springer, Singapore, pp 463–480
- Burnett AJ, Enyeart Smith TM, Wessel MT (2016) Use of the Social Cognitive Theory to Frame University Students' Perceptions of Cheating. *J Academic Ethics* 14(1):49–69. doi:10.1007/s10805-015-9252-4
- Callison-Burch C, Flournoy RS (2001) A program for automatically selecting the best output from multiple machine translation engines, Paper presented at the Proceedings of the Machine Translation Summit VIII
- Carroll J, Appleton J (2005). Towards consistent penalty decisions for breaches of academic regulations in one UK university. *Int J Educ Integr* 1(1):1–11
- Carter D, Inkpen D (2012) Searching for poor quality machine translated text : Learning the difference between human writing and machine translations, Paper presented at the Advances in Artificial Intelligence: 25th Canadian Conference on Artificial Intelligence, Canadian AI 2012, 28–30 May 2012, Toronto, Ontario, Canada
- Chen J, Kuznetsova P, Warren D, and Choi Y (2015). Déjà image-captions: A corpus of expressive descriptions in repetition. Paper presented at the Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

- Chen MH, Huang ST, Chang JS, Liou HC (2015b) Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Comput Assist Lang Learn* 28(1):22–40. doi:10.1080/09588221.2013.783873
- Clarke R, and Lancaster T (2013). Commercial aspects of contract cheating. Paper presented at the Proceedings of the 18th ACM Conference on Innovation and Technology in Computer Science Education.
- Cumming A, Lai C, Cho H (2016) Students' writing from sources for academic purposes: A synthesis of recent research. *J Engl Acad Purp* 23:47–58, <http://dx.doi.org/10.1016/j.jeap.2016.06.002>
- Crisp G (2007). Staff attitudes to dealing with plagiarism issues: Perspectives from one Australian university. *Int J Educ Integr* 3(1):3–15
- Dahl S (2007). Turnitin®: The student perspective on using plagiarism detection software. *Act Learn Higher*
- Davis M, Morley J (2015) Phrasal intertextuality: The responses of academics from different disciplines to students' re-use of phrases. *J Second Lang Writ* 28:20–35, <http://dx.doi.org/10.1016/j.jslw.2015.02.004>
- Evans C (2013) Making sense of assessment feedback in higher education. *Rev Educ Res* 83(1):70–120. doi:10.3102/0034654312474350
- Evering LC, Moorman G (2012) Rethinking plagiarism in the digital age. *J Adolesc Adult Lit* 56(1):35–44. doi:10.1002/JAAL.00100
- Fillenbaum S (1970) A note on the "Search after meaning": Sensibleness of paraphrases of well formed and malformed expressions. *Psychon Sci* 18(2):67–68. doi:10.3758/bf03335699
- Granitz N (2007) Applying ethical theories: Interpreting and responding to student plagiarism. *J Bus Ethics* 72(3):293–306. doi:10.1007/s10551-006-9171-9
- Hirvela A, Du Q (2013) "Why am I paraphrasing?" Undergraduate ESL writers' engagement with source-based academic writing and reading. *J Engl Acad Purp* 12(2):87–98, <http://dx.doi.org/10.1016/j.jeap.2012.11.005>
- Howard RM (1995) Plagiarisms, authorships, and the academic death penalty. *Coll Engl* 57(7):788–806. doi:10.2307/378403
- Inaba R, Murakami Y, Nadamoto A, and Ishida T (2007). Multilingual communication support using the language grid Intercultural Collaboration. Springer, Berlin Heidelberg, p 118–132
- Keck C (2006) The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *J Second Lang Writ* 15(4):261–278
- Keck C (2010). How do university students attempt to avoid plagiarism? A grammatical analysis of undergraduate paraphrasing strategies. *Writing & Pedagogy* 2(2):192–222
- Keck C (2014) Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *J Second Lang Writ* 25:4–22, <http://dx.doi.org/10.1016/j.jslw.2014.05.005>
- Kiros R, Zhu Y, Salakhutdinov R, Zemel RS, Torralba A, Urtasun R, Fidler S (2015) Skip-Thought Vectors, Paper presented at the Neural Information Processing Systems 2015, Montreal, Canada
- Lancaster T, Clarke R (2009) Automated essay spinning—an initial investigation, Paper presented at the 10 th Annual Conference of the Subject Centre for Information and Computer Sciences
- Lee A (2016) Student perspectives on plagiarism. In: Bretag T (ed) *Handbook of Academic Integrity*. Springer, Singapore, pp 519–535
- Li Y, Casanave CP (2012) Two first-year students' strategies for writing from sources: Patchwriting or plagiarism? *J Second Lang Writ* 21(2):165–180, <http://dx.doi.org/10.1016/j.jslw.2012.03.002>
- Lopez A (2008) Statistical machine translation. *ACM Computing Survey* 40(3):1–49. doi:10.1145/1380584.1380586
- Madera Q, García-Valdez M, and Mancilla A (2014). Ad text optimization using interactive evolutionary computation techniques Recent Advances on Hybrid Approaches for Designing Intelligent Systems. Springer, Heidelberg, pp 671–680
- McCarthy G (2014). Coaching and mentoring for business: Sage, London
- McCarthy G, Rogerson AM (2009) Links are not enough: using originality reports to improve academic standards, compliance and learning outcomes among postgraduate students. *Int J Educ Integr* 5(2):47–57
- Menai MEB (2012) Detection of plagiarism in Arabic documents. *Int J Inf Technol Comput Sci* 4(10):80–89
- Niño A (2009) Machine translation in foreign language learning: language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL* 21(02):241–258
- Patil S, Karekatti T (2015) The use of communication strategies in oral communicative situations by engineering students. *Language in India* 15(3):214–238
- Pfannenstiel AN (2010) Digital literacies and academic integrity. *Int J Educ Integr* 6(2):41–49
- Rabab'ah G (2016) The effect of communication strategy training on the development of EFL learners' strategic competence and oral communicative ability. *J Psycholinguist Res* 45(3):625–651. doi:10.1007/s10936-015-9365-3
- Rogerson AM (2014). Detecting the work of essay mills and file swapping sites: some clues they leave behind. Paper presented at the 6th International Integrity and Plagiarism Conference Newcastle-on-Tyne.
- Rogerson AM, Bassanta G (2016) Peer-to-peer file sharing and academic integrity in the Internet age. In: Bretag T (ed) *Handbook of Academic Integrity*. Springer, Singapore, pp 273–285
- Roig M (2001) Plagiarism and paraphrasing criteria of college and university professors. *Ethics & Behavior* 11(3):307–323
- Roig M (2016) Recycling our own work in the digital age. In: Bretag T (ed) *Handbook of Academic Integrity*. Springer, Singapore, pp 655–669
- Sambell K, McDowell L, Montgomery C (2013) Assessment for learning in higher education. Routledge, Abingdon, Oxon
- Scanlon PM, Neumann DR (2002) Internet plagiarism among college students. *J Coll Stud Dev* 43(3):374–385
- Shi L (2012) Rewriting and paraphrasing source texts in second language writing. *J Second Lang Writ* 21:134–148. doi:10.1016/j.jslw.2012.03.003
- Socher R, Huang EH, Pennin J, Manning CD, Ng AY (2011) Dynamic pooling and unfolding recursive autoencoders for paraphrase detection, Paper presented at the Advances in Neural Information Processing Systems (NIPS), Granada, Spain
- Somers H (2012). Computer-assisted language learning and machine translation. *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd, Hoboken, p 1–9
- Somers H, Gaspari F, Niño A (2006) Detecting inappropriate use of free online machine-translation by language students-A special case of plagiarism detection, Paper presented at the 11th Annual Conference of the European Association for Machine Translation–Proceedings, Oslo, Norway

- Spievak ER, Hayes-Bahanan P (2016) Creating order: The role of heuristics in website selection. *Internet Reference Services Quarterly* 21(1–2):23–46. doi:10.1080/10875301.2016.1149541
- Stamatatos E (2011) Plagiarism and authorship analysis: Introduction to the special issue. *Lang Resour Eval* 45(1):1–4. doi:10.1007/s10579-011-9136-1
- Suchan J (2014) Toward an understanding of Arabic persuasion: A western perspective. *Int J Bus Commun* 51(3):279–303. doi:10.1177/2329488414525401
- Sun Y-C (2012) Does text readability matter? A study of paraphrasing and plagiarism in English as a foreign language writing context. *Asia-Pacific Education Researcher* 21(2):296–306
- Thatcher SG (2008) China's copyright dilemma. *Learned Publishing* 21(4):278–284
- Walker J (1998) Student plagiarism in universities: What are we doing about it? *Higher Education Research & Development* 17(1):89–106
- Warnock S (2006) "Awesome job"—Or was it? The "many eyes" of asynchronous writing environments and the implications on plagiarism. *Plagiarism: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification* 1:178–190
- Word salad (2016). English Oxford Living Dictionaries online. Retrieved from [https://en.oxforddictionaries.com/definition/word\\_salad](https://en.oxforddictionaries.com/definition/word_salad). Accessed 23 Aug 2016

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---