

REVIEW

Open Access



What are Automated Paraphrasing Tools and how do we address them? A review of a growing threat to academic integrity

Jasper Roe^{1*}  and Mike Perkins²

*Correspondence:
Jasper.Roe@jcu.edu.au

¹ James Cook University
Singapore, Singapore, Singapore
² British University Vietnam, Hung
Yen, Vietnam

Abstract

This article reviews the literature surrounding the growing use of Automated Paraphrasing Tools (APTs) as a threat to educational integrity. In academia there is a technological arms-race occurring between the development of tools and techniques which facilitate violations of the principles of educational integrity, including text-based plagiarism, and methods for identifying such behaviors. APTs are part of this race, as they are a rapidly developing technology which can help writers transform words, phrases, and entire sentences and paragraphs at the click of a button. This article seeks to review the literature surrounding the history of APT use and the current understanding of APTs placed in the broader context of the educational integrity-technology arms race.

Keywords: Automated paraphrasing tools, Paraphrasing plagiarism, Technology-assisted plagiarism, Educational integrity

Introduction: defining the educational integrity-technology arms race

Universities are going through a period of unprecedented disruption, and concerns regarding breaches of academic integrity can be seen as part of the wider context of social, economic, and technological changes in higher education (Bretag et al., 2018). Access to online resources, accelerating Internet connection speeds, and global interconnectedness continue to progress, and while this has several positive benefits for academic work (including the dissemination of ideas and access to resources (Rogerson, 2020)), it brings with it more technologically advanced methods of committing academic misconduct and defying the norms, rules, and principles of educational integrity. This includes not only text-based plagiarism as we describe in the case of APTs, but also the availability of contract cheating services available through the 'booming' online sharing economy (Bretag et al., 2018). When describing plagiarism, we view it as part of the broader category of Academic Dishonesty (AD), defined by the International Center for Academic Integrity (2022) as a group of behaviors including plagiarism, cheating, lying, and deception. Academically dishonest behaviors often constitute academic misconduct,



© The Author(s) 2022, corrected publication 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

defined as engaging in fraud or deception through misrepresentation of work (Prescott, 1989).

While opportunities to engage in technologically assisted academic misconduct are growing, so are tools to assist in their detection. The development of these has become an active field of investigation in computer science and Natural Language Processing (NLP). This process is similar in nature to the concept of a military arms-race; with a pattern of competing development and acquisition of ever-stronger tools to evade and attack. As one method, software, or system is developed for engaging in breaches of educational integrity, a technological solution is shortly in development thereafter to combat it. Evidence of this can be seen through the work of Foltynek, Meuschke, & Gipp (2019), who found that between 2013 and 2018, 239 studies in the field of NLP focused on using technological means to identify complex forms of academic plagiarism. Some of these show great promise, with one tool developed by Foltynek et al. (2020) demonstrating accuracy of up to 99% in identifying machine-translated paraphrased text documents.

For each of these success stories, a new way of violating principles of educational integrity can equally be described. Alvi, Stevenson & Clough (2017) for example, highlighted how the use of homoglyphs can be employed by writers to replace letters with visually identical letters from other scripts, thus bypassing traditional text-matching anti-plagiarism software. Plagiarism in English using non-English source material is another important area of study. This has driven research in the identification of similar semantic meaning of two segments of text in different languages (Ferrero et al., 2017) to help detect when writers are taking existing text or ideas from non-English sources, translating it to English and claiming it as their own. As more of these techniques for engaging in violations of educational integrity appear, with them comes confusion and ambiguity. The lines between acceptable and unacceptable academic behavior are not universal, nor are they clear-cut. Rather, these behaviors exist on a continuum, and the place on the continuum that some new tools occupy is not entirely clear.

In this article, we aim to contribute to solving this problem through engaging in a detailed literature review of a category of tool that may be used to commit academic misconduct by aiding text-based plagiarism, that of Automated Paraphrasing Tools (APTs). We begin by describing the origins of APTs and their use in academic work. We then explore the relationship between language proficiency and APT use, and how APTs may or may not be used for in an academically dishonest way, referring to case studies from Dinneen (2021) and Prentice and Kinden (2018). Finally, we propose solutions and relevant limitations to tackling the problem of APTs in academia, as well as areas for future research.

Defining APTs and understanding their origins

Rogerson and McCarthy (2017) provide the clearest introduction and definition of what an APT is and does, stating that they are often web-based applications which use Machine Translation (MT) to transform one text into another, including between languages. MT varies in its level of sophistication and efficacy but is improving with advances in technology in the field of Natural Language Processing (NLP) and machine learning, although mistakes in output are still common (Rogerson, 2020). APTs were originally conceived to engage in 'text-spinning' as a method of achieving search engine optimization (Zhang, Wang, &

Voelker, 2014), and paraphrase in this field is required as originality is a key criterion for search engine optimization (Rogerson, 2020).

From this beginning in website development, APTs have found a second user-base in academia, allowing writers to disguise source material in the submission of assessments (Rogerson, 2020) and bypass plagiarism detection services which use text-matching algorithms. The underlying factors leading to the use of these tools is not well understood. The relationship between language proficiency and plagiarism may lead to the conclusion that APT users are primarily novice students who are not native English speakers, but are instead using English as a Foreign Language (EFL) (Rogerson & McCarthy, 2017). However, Rogerson (2020) also argues that professional scholars and researchers may equally make use of these tools. To demonstrate Rogerson's (2020) point, Ansorge, Ansorgeova, and Sixsmith (2021) described a single case of an article published in a journal which was later found to be likely to have used an APT. The authors used an online tool called 'DiffChecker' to identify 817 unique differences between the suspected source text (another journal article) and the published text; the tool found that it was highly likely the second text was produced by a machine, suggesting the use of an APT.

The relationship between APT use, paraphrase plagiarism, and language proficiency

Although the rules and norms of acceptability may vary between institutions and contexts, students in Higher Education must follow principles of academic integrity, which are built on values of honesty, fairness, trust, respect, and responsibility (Lynch, Salmonson & Glew et al., 2021). One of the methods by which students are expected to show these values is through paraphrasing: a skill which demonstrates that they can understand works that they have read, and distil, reproduce, comment on, or critique these ideas while maintaining proper acknowledgement of sources (Rogerson & McCarthy, 2017). Inappropriate paraphrase on the other hand, may contain the same lexis and overall structure as the original source material (Oshima & Hogue, 1999), thus resulting in plagiarism in some cases. Paraphrasing is a critical skill for successful writing, but can be difficult for students, especially for those who are not writing in their first language (Chen et al., 2015; Rogerson, 2017; Shi, 2012). This is one important factor in understanding the relationship between language proficiency and the use of APTs.

Non-native English writers were found by Keck (2006) to use more 'near copies' of phrases than native English-speaking writers, and the relationship between language proficiency and ability to paraphrase has also been shown as related to the level of students' text comprehension (Erhel & Jamet, 2006). Insufficient knowledge may also lead to students being unable to think of a way to restate an idea (Rogerson & McCarthy, 2017). Therefore, a lower level of ability in English may lead to lower text comprehension, resulting in poorer paraphrasing. Several studies have equally found a negative association between English proficiency and engaging in plagiarism, such as Bretag (2007), Li (2015), Pennycook (1996), Marshall and Garry (2006), Perkins, Gezgin and Roe (2018), and Chen and Ku (2007). However, Keck (2014) also found that novice writers have also been shown to rely more heavily on copying from source material, so experience may also play a role in the ability to paraphrase.

One further complicating factor when understanding APT use and its role as an academically dishonest behavior is a lack of clarity as to what constitutes appropriate and inappropriate paraphrasing. Sun and Yang (2015) state that the definition of plagiarism and paraphrasing in academic work is unclear, leading to a lack of consensus. Shi (2004) proposes that paraphrasing be considered as matching more than two to three words from the original source material, while others state that even the duplication of words can be an indicator of plagiarism (Benos et al., 2005). Sun (2013) points out that with the varying requirements of different disciplines in academia, what is and is not acceptable may also vary. The lack of consensus on what constitutes appropriate paraphrasing may be one factor that affects students' ability in academic writing and makes it more difficult to understand the use of APTs and to what extent they constitute academic misconduct. By reviewing the types of APT and how they are used however, a clearer perspective on when APT use constitutes AD can be formulated.

Types of APTs and their use in academic work

There are several different varieties of APTs, and all are not created equal. Prentice and Kinden (2018) highlight that between Rogerson and McCarthy's (2017) initial finding of 550,000 results from a search engine query for paraphrasing tools, the number of results had reached 3 million by 2018. A search for this term in November 2021 obtained results of approximately 4.5 million; highlighting not only the growing number of APTs available, but also the increased interest in this field shown by both scholars and the general public alike. Close inspection of some of the top-ranking results on search engines shows that some APT applications seem to be mirror-duplicates of the same framework and technology which are free to use and rely on advertisements Others offer a greater range of fee-based subscription services, including alterable parameters of replacement at the lexis, phrase, or sentence level (Prentice and Kinden, 2018). This suggests that there may be large gaps between the efficacy, accuracy, and sophistication of the APTs which are presently being used.

One other variety of APTs are those which are used for pedagogical purposes and do not constitute a violation of principles of educational integrity. In the field of EFL, these can be indispensable tools for teaching paraphrasing as a skill. Chen et al. (2015) for example, demonstrated success in creating a corpus-based tool to suggest paraphrases using a parallel Chinese-English corpus, and found that 90% of the sample ($N = 55$) preferred to write using their assistive paraphrasing tool, and 75% felt that the tool benefited their writing. This demonstrates that for students who are practicing English writing as English as a Foreign Language (EFL), such APTs can be a valuable resource for learning. That said, if learners come into contact with these APTs and they are not properly contextualized by the instructor, they have the potential to cause confusion as to what is and what is not acceptable for formal assessments. This is compounded by the common use of corpora and paraphrasing tools in the English language classroom, something that many English as a Foreign Language speakers may experience. If an EFL student is introduced to an APT by a teacher, for example in a university English class environment, it follows that they may find it confusing if it is deemed unacceptable for use in an assessment and results in them subsequently being accused of plagiarism.

In terms of how APTs are used (except for pedagogical APTs) both free and paid varieties tend to follow a similar system. Users input raw text into an interface, press an action button, and then retrieve the automatically generated output, which in theory, encodes and communicates the same core ideas or message as a different set of words. However, given the variable effectiveness of MT, this can result in the production of incomprehensible text, which has been referred to as ‘word salad’ (Rogerson & McCarthy, 2017). As an example, Prentice and Kinden (2018) found that in the discipline of health sciences, the use of paraphrasing tools resulted in medical terminology being substituted for incomprehensible words that lacked meaning. This can be one of the clear indications that an APT has been used.

In terms of how users engage with APTs, following the authors’ experiences, a general set of strategies for their illicit use in academic writing can be outlined as follows. Users first locate texts which are relevant to the subject at hand, and then copy material verbatim from the source material, (commonly websites, textbooks and journal articles) and enter it into the tool. Students may also engage in ‘back translation’ (Jones, 2009; Dinneen, 2021) in which they copy the original source material, translate it into a foreign language (again using a MT tool such as Google Translate) and then translate it back to English, resulting in a paraphrased version of the original. Users may then pass this through an APT again, in a 3-step process. By doing so, the writer may believe they are able to bypass plagiarism detection software, reduce the amount of effort required to produce original text through paraphrase, or may simply feel that they have successfully engaged in paraphrasing, thus not committing any violation. If a ‘word salad’ (Rogerson & McCarthy, 2017) is produced where text is incoherent, writers may attempt to proof-read and edit the paraphrased text to increase readability and avoid suspicion. These uses constitute Academic Dishonesty and are in our view paraphrasing plagiarism.

A review of APT case reports and the risks presented

While we have made clear which cases we argue constitute legitimate (pedagogical) uses of APTs and which constitute AD and paraphrasing plagiarism, this may not be clear to students who intend to use an APT. Sun (2013) discusses the possible generational-cultural dimensions that may affect use, quoting Weiler (2005) argument that for some generations of learners, learning focuses on seeking rather than critiquing information, meaning that learners may not see why text reproduction is academic misconduct. Students may then not clearly understand why APT use can result in plagiarism. Evidence for this comes from Bowen and Nani’s (2021) findings that Thai students were uncertain about the difference between patchwriting; a simplistic form of superficial (Rogerson & McCarthy, 2017) or close paraphrasing (Keck, 2010) and acceptable paraphrasing.

One example of such seemingly unintentional use of an APT to commit paraphrasing plagiarism is given by Prentice and Kinden (2018), who describe a situation of a student using an APT to paraphrase text from file-sharing sites, while providing the original source in a reference list. Although the inclusion of the original source material in a reference list implies that the student did not intend to deceive, this can under most definitions be considered plagiarism. On the other hand, an EFL student writing in their first language, and then translating it to English, followed by passing it through an APT, may be considered poor academic practice, or a disingenuous representation of their

own abilities, but not, by definition, plagiarism. This is a debatable example, given that the answer to whether the text is in the students' own words is not clear cut. Some may argue that the student's ideas were initially created by the student, and only the phrasing and linguistic medium has been changed, where others may state that the student has not met the requirements of writing in the target language and has attempted to deceive the assessor that they have done so, constituting Academic Dishonesty and paraphrasing plagiarism.

A further case that may create debate is a report from Dinneen (2021), who describes a student who had copied 75% of the submitted text for an assessment but remained convinced that as they had used in-text citations, and changed the wording of the authors' original text (through using an APT), they had not committed any form of misconduct nor plagiarized. Based on the interpretation of the institution's plagiarism policy, it was found that there was no indication that algorithm-driven paraphrase constituted academic misconduct, meaning that in essence the student was correct (Dinneen, 2021). Our position on this is that despite not meeting the technical definition of academic misconduct based on the institution's lack of policy, this does not change the core fact that the student's submitted work was not their own. While some institutions may already have implemented policies to counteract these kinds of cases, the case study highlights the need for universal adoption of guidelines for institutions to deal with APT usage as it becomes more widespread.

While then, there are many areas of debate surrounding APT use, the fact remains that they are a serious and current threat to academic integrity, which can hide plagiarism and help to facilitate collusion (Wahle et al., 2021). APTs can serve to reduce the ability of text-matching software used to help identify potential cases of plagiarism, thus weakening one of the most effective current diagnostic tools for academic misconduct and plagiarism (Wahle et al., 2021). These tools not only represent a risk for students at the undergraduate and postgraduate level, but even for faculty and researchers who may wish to expand their output through publishing paraphrased versions of the same work while adding no new content. Rogerson (2020) highlights other risks, given that there is no publicly available information on how much data is collected from these tools, and what happens to this stored data. In all, this paints a concerning picture for APT use in academia.

Addressing APTs: What's next?

Given the lack of consensus on several key issues relating to APTs, the question of how institutions and educators should address these tools is complex. Several strategies are available to help combat the use of APTs at present, but all carry some limitations, especially as more is found out about how these tools are used in practice, and as these tools continue to evolve.

Under the arms-race scenario, institutions and educators may look towards developments in technology for identifying the use of APTs. Current options in development include Longformer, which attempts to identify machine-based plagiarism, and DSpin, created by Zhang, Wang, and Voelker (2014), which aims to automatically identify text created by APTs. Foltynnek et al. (2019)'s systematic literature review of computational methods of plagiarism detection notes that there have been large improvements

in technological solutions to identifying plagiarism, which are mainly the result of improved methods of semantic analysis, as well as the use of non-textual elements of written work and the use of machine learning. This means that with the continued development of the field, the ability of software to identify the use of APTs and other difficult to detect, or 'complex plagiarism' (Perkins, Gezgin & Gordon, 2019), may be on the horizon. Other authors, such as Perkins, Gezgin and Roe (2020) also highlight that while current software is not yet able to accurately identify these more complex cases of plagiarism, emerging fields of deep learning and neural network technologies have high potential in easing academic misconduct issues in higher education in future.

Whether an automated tool will be usable to detect APTs on a highly accurate, accessible basis in future is still then, an unknown, but machine-translated text is usually identifiable by an individual reading the material (Carter & Inkpen, 2012). In terms of the arms-race metaphor however, it may not be long before proficient speakers start to find it more challenging to distinguish between APT text and human-written text, as APTs continue to develop. This leads us to advocate for one established method that supersedes the arms race: training. Training is important, as at present, despite advances in technology, identifying plagiarism remains a social activity that currently requires human intervention in identification (Weber-Wulff, 2019). It is well established that training both students and faculty can have a positive effect on reducing breaches of academic integrity. Duff et al. (2006) found that over a three-year period, providing cross-cultural training on critical scholarship in the Western academic tradition, and taking an approach towards guiding students rather than focusing on detection and punishment led to improvements in scholarship. Dawson, Sutherland-Smith and Rickson (2020) found that faculty using Turnitin's Authorship Investigate tool led to significant increases in their ability to detect contract cheating, and Dawson and Sutherland-Smith (2019) demonstrated that marker training is helpful in identifying contract cheating. Perkins, Gezgin and Roe (2020) identify how academic misconduct education and training of students can potentially lead to a reduction in the instances of plagiarism that take place, and Du (2019) found that a single six-hour period of instruction reduced plagiarism in participants writing. Recognizing the broader reasons which may lead to plagiarism, and accounting for this in the development of supportive academic policies and practices is therefore of importance in reducing the usage of these tools amongst students. Martin (2004) also states that a policy of effective training, modeling, and rewards, is more effective than a disciplinary approach to poor practice. It is important to note that cultural norms should not be ignored in implementing such training, as the Western notion of academic integrity is not universal, and has been implicated as dismissive of other cultures, in particular the Eastern academic tradition of duplication as homage (Stowers & Hummel, 2011; Roe & Perkins 2020). To take a student-centered approach then, would mean to continue providing students with greater training on what these tools are, how they can be used legitimately, and how illicit use can be avoided.

However, if student training is to be used as an initial proactive approach to dealing with APTs, then a clear communication strategy should be devised to ensure that students understand the difference between the use of such tools pedagogically in the English as a Foreign Language (EFL) language classroom (Chen et al., 2015), and their use individually to produce assessed work in their disciplines of study. Training for both

students and faculty should include examples of the resulting ‘word salads’ (Rogerson & McCarthy, 2017) and poorly paraphrased sentences to emphasize the potential risks of the software producing unsatisfactory work, including typical features such as unclear sentence meaning, missing data, and incorrect referencing (Ansorge, Ansorgeova, & Sixsmith, 2021), aside from the serious risk of violating principles of educational integrity, as recommended by Nino (2009). This avoids the situation in which educators are forced to make difficult decisions without adequate training and recognizes that academics play a vital role in the detection of academic misconduct (Bretag & Mahmud, 2009).

Conclusion

As technology continues to accelerate, the rate of development in advanced tools which manipulate language for a variety of purposes, including to aid academic work both legitimately and illicitly, will continue to grow. The role of academics is to decipher their use, understand why and how they are used, and make judgements on at what point this constitutes an unacceptable usage. As Dinneen (2021) states, there is currently a ‘silence’ on the appropriate use of digital tools in institutional academic integrity policies. This article has sought to remedy this through the review of current literature pertaining to APTs and offer insight into issues which institutions and faculty might face when confronted with this growing threat among both native English speaking and EFL students. We have also identified that the current approach of combating the illicit use of APTs through the development of technical solutions is promising but may continue to form an arms-race scenario. We therefore advocate for training as the most important tool in both reducing the use of APTs by students, as well as improving the ability of faculty to detect any such use. Finally, as recommended by Rogerson (2020) additional investigations should aim to develop broader social insights into the use of APTs. Further research into the effectiveness and structure of APTs, as well as why students use them, will further illuminate this challenging topic.

Abbreviations

APTs	Automated Paraphrasing Tools are software applications which produce paraphrased text through user input
SEO	Search Engine Optimization is the process of a website obtaining a higher ranking on a search engine to enjoy greater visibility
EFL	English as a Foreign Language is the speaking of English as a language other than one’s own mother tongue
NLP	Natural Language Processing is an emerging field involving artificial intelligence, linguistics, and machine learning

Acknowledgements

The authors make no acknowledgements.

Authors’ contributions

Mr Jasper Roe contributed the majority of the review and writing. Dr Mike Perkins contributed significantly to the review, writing, editing and finalization of the manuscript. The author(s) read and approved the final manuscript.

Funding

The authors received no funding for this research.

Availability of data and materials

No data is made available from this article.

Declarations

Competing interests

The authors declare no competing interests.

Received: 9 December 2021 Accepted: 17 May 2022

Published online: 07 July 2022

References

- Alvi F, Stevenson M, Clough P (2017) Plagiarism Detection in Texts Obfuscated with Homoglyphs. In: Jose J et al (eds) *Advances in Information Retrieval. ECIR 2017. Lecture notes in computer science*, vol 10193. Springer, Cham. https://doi.org/10.1007/978-3-319-56608-5_64
- Ansorge L, Ansorgeová K, Sixsmith M (2021) Plagiarism through paraphrasing tools—the story of one plagiarized text. *Publications* 9(4):48
- Bailey J. (2018) A brief history of Article Spinning. Plagiarismtoday.com. Retrieved August 20, 2021: www.plagiarismtoday.com/2018/03/08/a-brief-history-of-article-spinning
- Benos DJ, Fabres J, Farmer J, Gutierrez JP, Hennessy K, Kosek D, Lee JH, Olteanu D, Russell T, Shaikh F, Wang K (2005) Ethics and scientific publication. *Adv Physiol Educ* 29(2):59–74. <https://doi.org/10.1152/advan.00056.2004>
- Bowen NEJA, Nanni A (2021) Piracy, playing the system, or poor policies? Perspectives on plagiarism in Thailand. *J Engl Acad Purp* 51. <https://doi.org/10.1016/j.jeap.2021.100992>
- Bretag T (2007) The emperor's new clothes: yes, there is a link between English language competence and academic standards. *People Place* 15(1):13–21
- Bretag T, Harper R, Burton M, Ellis C, Newton P, Rozenberg P, Saddiqui S, van Haeringen K (2018) Contract cheating: a survey of Australian university students. *Stud High Educ* 44(11):1837–1856
- Bretag T, Mahmud S (2009) Self-plagiarism or appropriate textual re-use? *J Acad Ethics* 7(3):193–205
- Carter D, Inkpen D (2012) Searching for Poor Quality Machine Translated Text: Learning the Difference between Human Writing and Machine Translations. In: Kosseim L, Inkpen D (eds) *Advances in Artificial Intelligence. Canadian AI 2012. Lecture notes in computer science*, 7310. Springer, Berlin
- Center for Academic Integrity. (n.d.). *The Fundamental Values of Academic Integrity*. <https://secure2.mc.duke.edu/academicintegrity/pdf/FVProject.pdf>
- Chen MH, Huang ST, Chang JS, Liou HC (2015) Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Comput Assist Lang Learn* 28(1):22–40
- Chen T, Ku NKT (2007) EFL students: factors contributing to online plagiarism. In: Roberts TS (ed) *Student plagiarism in an online world: problems and solutions*. IGI Global, New York, pp 77–91
- Dawson P, Sutherland-Smith W (2019) Can training improve marker accuracy at detecting contract cheating? A multi-disciplinary pre-post study. *Assess Eval High Educ* 44(5):715–725. <https://doi.org/10.1080/02602938.2018.1531109>
- Dawson P, Sutherland-Smith W, Rickson M (2020) Can software improve marker accuracy at detecting contract cheating? A pilot study of the Turnitin authorship investigate alpha. *Assess Eval High Educ* 45(4):473–482. <https://doi.org/10.1080/02602938.2019.1662884>
- Dey K, Shrivastava R, Kaushik S (2016) A paraphrase and semantic similarity detection system for user generated short-text content on microblogs. In: *Proceedings of COLING 2016, 26th International Conference On Computational Linguistics: Technical Papers*:2880–2890
- Dinneen C (2021) Students' use of digital translation and paraphrasing tools in written assignments on direct entry English programs. *Engl Aust J* 37(1):40–54
- Duff AH, Rogers DP, Harris MB (2006) International engineering students: avoiding plagiarism through understanding the Western academic context of scholarship. *Eur J Eng Educ* 31(6):673–681. <https://doi.org/10.1080/03043790600911753>
- Du Y (2019) Evaluation of intervention on Chinese graduate students' understanding of textual plagiarism and skills at source referencing. *Assessment & Evaluation in Higher Education*
- Erhel S, Jamet E (2006) Using pop-up windows to improve multimedia learning. *J Comp Assist Learn* 22(2):137–147
- Ferrero J, Agnes F, Besacier L, Schwab D (2017) Using word embedding for cross-language plagiarism detection. *arXiv preprint arXiv:1702.03082*
- Foltynek T, Meuschke B, Gipp B (2019) Academic plagiarism detection: a systematic literature review. *ACM Comput Surv* 52(6):1–42
- Foltynek T, Ruas T, Scharpf P, Meuschke N, Schubotz M, Grosky W, Gipp B (2020) Detecting machine-obfuscated plagiarism. In: *International conference on information*. Springer, Cham, pp 816–827
- Franco-Salvador M, Gupta P, Rosso P, Banchs RE (2016) Cross-language plagiarism detection over continuous-space- and knowledge graph-based representations of language. *Knowl Based Syst* 1(111):87–99
- Guerrero-Dib JG, Portales L, Heredia-Escorza Y (2020) Impact of academic integrity on workplace ethical behavior. *Int J Educ Integr* 16(2):1–18. <https://doi.org/10.1007/s40979-020-0051-3>
- Jalilifar A, Soltani P, Shooshtari ZG (2018) Improper textual borrowing practices: evidence from Iranian applied linguistics journal articles. *J Engl Acad Purp* 35(1):42–55
- Jones M (2009) Back-translation: the latest form of plagiarism. *The 4th Asia Pacific conference on educational integrity*:1–7 Wollongong, Australia
- Keck C (2006) The use of paraphrase in summary writing: a comparison of L1 and L2 writers. *J Second Lang Writ* 15:261–278
- Keck C (2010) How do University students attempt to avoid plagiarism? A grammatical analysis of undergraduate paraphrasing strategies. *Writ Pedagogy* 2(2):193–222. <https://doi.org/10.1558/wap.v2i2.193>
- 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
- Lester MC, Diekhoff GM (2002) A comparison of traditional and internet cheaters. *J Coll Stud Dev* 43(6):906–911
- Li Y (2015) Academic staff's perspectives upon student plagiarism: a case study at a university in Hong Kong. *Asia Pacific J Educ* 35(1):14–26

- Lynch J, Salamonson Y, Glew P, Ramjan L (2021) "I'm not an investigator and I'm not a police officer" - a faculty's view on academic integrity in an undergraduate nursing degree. *Int J Educ Integr* 17(19). <https://doi.org/10.1007/s40979-021-00086-6>
- Marshall S, Garry M (2006) NESB and ESB students' attitudes and perceptions of plagiarism. *Int J Educ Integr* 2(1):26–37
- Martin B. (2004). Plagiarism: Policy against cheating or policy for learning? Available at: <http://www.bmartin.cc/pubs/04plag.pdf>. Accessed 29 Nov 2021
- Meuschke N, Gipp B (2013) State of the art in detecting academic plagiarism. *Int J Educ Integr* 9(1):50–71
- Niño A (2009) Machine translation in foreign language learning: language learners' and tutors' perceptions of its advantages and disadvantages. *Recall* 21(2):241–258
- Oshima A, Hogue A (1999) *Writing academic English*. Addison Wesley Longman, New York
- Pennycook A (1996) Borrowing Others' words: text, ownership, memory, and plagiarism. *TESOL Q* 30(2):201–230
- Perkins M, Gezgin UB, Gordon RD (2019) Plagiarism in higher education: classification, causes and controls. *Pan-Pac Manag Sci* 2:3–21. Available at: https://www.researchgate.net/publication/354143709_Plagiarism_in_higher_education_classification_causes_and_controls.
- Perkins M, Gezgin UB, Roe J (2018) Understanding the relationship between language ability and plagiarism in non-native English speaking business students. *J Acad Ethics* 16(4):317–328. <https://doi.org/10.1007/s10805-018-9311-8>.
- Perkins M, Gezgin UB, Roe J (2020) Reducing plagiarism through academic misconduct education. *Int J Educ Integr* 16(3). <https://doi.org/10.1007/s40979-020-00052-8>
- Peters M (2019) Academic integrity: an interview with Tracey Bretag. *Educ Phil Theory* 51(8):751–756
- Prentice FM, Kinden CE (2018) Paraphrasing tools, language translation tools and plagiarism: an exploratory study. *Int J Educ Integr* 14(1):1–16
- Prescott PA (1989) Academic misconduct: considerations for educational administrators. *J Prof Nurs* 5(5):283–287
- Roe J, Perkins M (2020) Learner autonomy in the Vietnamese context: a literature review. *Asian J Univ Educ* 16(1):13–21. <https://doi.org/10.24191/ajue.v16i1.8490>
- Rogerson AM (2014) Detecting the work of essay mills and file swapping sites: some clues they leave behind. *Sydney Business School-Papers* 1(1):1–9
- Rogerson AM (2017) Detecting contract cheating in essay and report submissions: process, patterns, clues and conversations. *Int J Educ Integr* 13(10):1–17
- Rogerson AM (2020) The use and misuse of online paraphrasing, editing and translation software. In: *A research agenda for academic integrity*. Edward Elgar Publishing, Cheltenham
- Rogerson AM, Mccarthy G (2017) Using internet based paraphrasing tools: original work, patchwriting or facilitated plagiarism? *Int J Educ Integr* 13(2):1–15
- Roig M (2010) Plagiarism and self-plagiarism: what every author should know. *Biochem Med* 20(3):295–300
- Shi L (2004) Textual borrowing in second-language writing. *Writ Commun* 21(1):171–200
- Shi L (2012) Rewriting and paraphrasing source texts in second language writing. *J Second Lang Writ* 21(2):134–148
- Stowers RH, Hummel JY (2011) The use of technology to combat plagiarism in business communication classes. *Bus Commun Q* 74(2):164–169
- Sun YC (2013) Do journal authors plagiarize? Using plagiarism detection software to uncover matching text across disciplines. *J English Acad Purp* 12(4):264–272
- Sun YC, Yang FY (2015) Uncovering published authors' text-borrowing practices: paraphrasing strategies, sources, and self-plagiarism. *J English Acad Purp* 20(1):224–236
- The International Center for Academic Integrity (2022) Core Values. <https://academicintegrity.org/resources/fundamental-values>. Accessed 1 Jan 2022
- Wahle J, Ruas T, Foltynnek T, Meuschke N, Gipp B (2021) Identifying machine based plagiarism. Arxiv preprint arxiv:2103.11909
- Weber-Wulff D (2019) Plagiarism detectors are a crutch, and a problem. *Nature* 567(7749):435–436
- Weiler A (2005) Information-seeking behavior in generation Y students: motivation, critical thinking, and learning theory. *J Acad Librariansh* 31(1):46–53
- Zhang Q, Wang DY, Voelker GM (2014) Dspin: detecting automatically spun content on the web. In: *Proceedings NDSS Symposium*, pp 23–26

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

