DE-EMPHASIZING PLAGIARISM WITH BIG DATA: MODERN APPROACHES AND ETHICAL CONSIDERATIONS

Daniel Dusza

Abstract: This article introduces present trends in Big Data analytics and how they might be implemented to de-emphasize plagiarism. Regrettably, academic institutions have continued to rely on essay-based coursework and written reports as a basis of assessment. As a result of the COVID-19 pandemic, there has been a mass migration into online learning, and with it, a further increase in the reliance on textual content. With more writers now composing online in the absence of face-to-face accountability to peers and teachers, the risks to academic integrity through plagiarism and contract cheating should be expected to increase.

Three empirical research studies were used to investigate how the writing process might be automatically and covertly monitored to measure the dynamics of compositions. The areas of interest include the equipment, the method of data management, and the information that could be gleaned from the recorded data. Each study is progressively more complex, and presented in a manner to support the future formulation of a framework for implementation into education.

The results indicate that the equipment that is readily available to most students is capable much more than composing and transmitting a written manuscript. Currently, technology is capable of identifying writing problems and providing assistance to help writers navigate even the most difficult tasks in composition. The technological solutions suggested in this paper provides far more than plagiarism detection. The results in this paper indicate that future writing will be supported through process verification, semantic network authentications, and other certifications that will form part of the future requirements of assessment and academic integrity.

Key words: plagiarism, TPACK, Big Data, AI, academic-integrity

Introduction

The 'postmodern' understanding of plagiarism extends beyond the mere analysis of textual data. Until now, the written product has been the central issue in academic integrity, and in detecting transgressions and foul play. However, over the past few decades the focus has shifted more onto the writer, their methods of where they glean information, how and what information is synthesized, and the integrity of the knowledge they are offering.

Three empirical writing research methods are presented in this article, which focus primarily on the writer during the writing process. This significant paradigm-shift away from the more conventional approach of analysing the written product is where the battle for academic integrity needs to be attentive. This paradigm shift takes the power away from ghost writers and essay mills, because the focus is removed from the product onto the writer, their ability, their behaviour, and their skills of using technology appropriately to author a final manuscript.

The first technique discusses a pedagogical approach. In this action research study (Dusza, 2020), three years of classroom data was collected to investigate the causes of

plagiarism in the writing process. The study involved EFL Japanese college students and specifically identified areas where the risk of plagiarism was most prevalent. In subsequent classes, online tools were offered to students to help scaffold their writing and language skills. At strategic times in the writing process, plagiarism detection software use was introduced. In fact, the use of the software became part of the assessment criteria. Additionally, assessment evolved to conform with Lancaster and Clarke's (2017) concept of mixed modalities, which was interpreted as assessing the process, more than the product. The results from the three-year action research study indicated that the primary reason students did not improve in quantity, quality and originality was because they were absent from strategic lessons, in which case, a grade could only be awarded according to the qualification that these less skilled writers deserve (Lancaster & Clarke, 2017, p. 216).

The second technique involves a study (Vasilopoulos, 2018), which used a collection of mixed data to examine the use of third party information (i.e., writing from the internet) in university students' essays. The feature of interest in the Vasilopoulos study is the covert use of technology to collect students' searching, reading, and writing behaviour. The study used readily available 'spy ware' to monitor web-page browsing, text manipulation, google searches and translations, in fact, everything that the students did on their devices. This use of this Big Data is presented as the foundation of authenticity certification, and should permeate into academic writing. This paper proposes that, in the future, ethical writers will use verification devices (i.e. embedded monitoring programs in machines or based on the Internet) to approve the integrity of authored material, throughout the complete writing process. Therefore, the integration of pedagogy mentioned previously, will become a requisite skill.

The logical progression from monitoring individuals and their devices, is to look at their interaction with peers and other non-structured communications. This level of "surveillance" verification collects information on the locations and positions of communication. The process involves monitoring the interaction between individuals and their networks through emails, social media, paper-based, and telecommunications. This third technique uses data analyses techniques as natural language processing (NLP). In the study feature in this paper, NLP data was collected from both internal and web based network activity, and includes social media, emails, and online meeting devices. The research (Kramer & Russell, 2014) provides a useful starting point to understand specifically how to extract communication and dynamic information intelligence from Emails. Even though, this level of processing may seem contemporary, it is already outdated. Similar processing can now yield results from data collected from audio, video, semantic analysis of communications, and other even more abstract forms of data.

The questions addressed in this paper are as follows.

1. Can the importance of plagiarism be deemphasized?
2. Can plagiarism be reduced by replacing human error, with technical solutions?
3. Can the burden of writing be reduced with smart technology?

4. Is it possible to certify the process behind the completed manuscript, so then, and only then, the value of the words on the paper and the message they contain be included in the academic body of knowledge?

While it is impossible to answer these questions in a single paper, the following sections will provide enough evidence to support the belief that we can revolutionise the writing process while reinforcing academic integrity, and thereby, de-emphasizing the influence of plagiarism.

Featured Research Review

In the recent Plagiarism Across Europe and Beyond online conference, Bretag (2020) outlined three constituents to research integrity that I found essential: first, clearly identify the field of research; second, identify relevant literature to support the theoretical and conceptual framework of the study, and third, utilise a methodology that is credible, or at least justifiable. The following is an honest attempt to provide information that is not only reputable, but will be useful in the future maintenance of academic integrity.

Pedagogical Issues

It has been long known that any good writing task is designed to counter attempts at plagiarism. However, little is ever presented to teachers regarding approaches to tackle the issues surrounding plagiarism. More importantly, little research has been available to identify when and why plagiarism presents the greatest risk. In the least, we should be looking at information on how students construct essays to inform us on how to make adjustments in the future (Todd, 2010).

Method

A pedagogical solution was in fact developed and studied over a three-year period (Dusza, 2020). This action research collected classroom data to investigate the causes of plagiarism within the writing process. The first phase of the study was to specifically identify where in the writing process was the greatest risk of plagiarism, and where it was most prevalent. It is widely accepted that the literature review and synthesis of this knowledge is where even experienced writers have the greatest burden. This burden is usually where students require the greatest levels of assistance, and it is assistance that is extremely time-consuming. Therefore, teachers are ill equipped to provide the necessary assistance students need. The intentional forms of plagiarism most evident in this stage are pseudo writing and plagiarising (see Vasilopoulos in the next section).

In the following semester, these problems were alleviated by providing training in the use of online tools, to help scaffold students' writing and language skills. At strategic points in the writing process, plagiarism detection software use was introduced, and students were given training in how to use other tools to reduce the risk. In fact, the use of software became a part of the assessment criteria. Additionally, the focus of assessment evolved to conform with Lancaster and Clarke's (2017) concept of mixed modalities, which was interpreted as assessing the process, more than the product.

Result

The results indicated that even though substantial time was required to introduce students to using technology, the amount of writing, and the quality of the writing had improved remarkably, albeit less than statistically significant (Dusza, 2020). The reason that some participants did not improve in quantity, quality, or originality was because they were absent from these strategic lessons. The research concluded that integrating technology and skills to deal with writing problems during the writing process, and assessing these skills during the writing process had the greatest impact on de-emphasizing plagiarism. Unfortunately, the few who could not attend consistently missed a valuable learning experience; they could only be awarded a grade based on an inferior finished product and not on the qualification that these less-skilled writers deserve (Lancaster & Clarke, 2017, p. 216).

Conclusion

The conclusion for the present study is that by integrating technology and pedagogy and using formative assessment, writing tasks can be effective in reducing the requirements that make plagiarism an issue (Todd, 2010).

Covert data collection on user devices

The previous section identified two areas in the writing process that present the biggest challenges to second-language writers, and therefore, the greatest risk of plagiarism. Synthesis of information and paraphrasing present a considerable burden to second language learners, and is also the area where the risk of plagiarism is presented for most writers. Understandably, the skills of summarising/synthesising, and paraphrasing are skills that are only developed with significant amounts of practice, even in our first language.

A research paper presented at the 2018 Symposium for Second Language Writing, inspired the pursuit of covert data collection presented in this paper. The research by Gene Vasilopoulos (2018) illustrated the astounding potential that the collection of covert data could provide to monitoring the writing process, and reporting plagiarism, and potentially the automated responses to counter it.

Vasilopoulos (2018) studied the intersection between L2 writing, plagiarism, and the influence of technology in the writing process, and the areas at greatest risk. While the study in the previous section was looking at pedagogical/technological solutions, Vasilopoulos was concerned with the mechanics of avoiding plagiarism (i.e. quotation marks, copy and paste plagiarism, translation without referencing, poor or incorrect referencing, and poor paraphrasing and summarising). She was also interested in measuring intent.

Method

The metric for measuring plagiarism involved identifying the following:

- *pseudo-writing* – similar to traditional plagiarism, but is void of physical rewriting by copying and pasting internet based text to form 'original' content in an attempt to reduce the [cognitive] workload of writing (Skaar, 2015),

- *de-plagiarising* – using a thesaurus or other applications to replace words (Stapleton, 2010), or copy-and-paste text, then 'cleanse' it to avoid detection (Wrigley, 2019),

- *paraphrase tools* – online tools that automatically change the words to mask copying, also known as 'spin-writers' (Dusza, 2019; Rogerson & McCarthy, 2017),

- *translation software* – which is, on occasion, used on entire articles (Jones & Sheridan, 2017), but is also useful in the writing development process (Wrobel, 2019), and

- *facilitated plagiarism* – actively seeking to use some form of easily accessible Internet-based sources to prepare or supplement submission material for assessment by others (Stamatatos, 2011).

The task required students to write a 1000-word essay that was both demanding contextually and assessed the skills of *research* from valid sources, correct *referencing* including citation, and *proofreading* for grammar, vocabulary, and referencing. Although these tasks were adequately performed by technology, the cognitive process was important to 'push' the participants into making decisions to incorporate the items mentioned in the above measurement metrics.

Collection and processing of Big Data

The intriguing aspect of this research was the collection and processing of the volumes of unstructured and structured data. By definition, Big Data is not just size, but a dynamic. It includes structure and unstructured data, highly variable in volume, and in transmission rates (Vaitsis, Hervatis, & Zary, 2016). The analysis conducted in this Vasilopoulos study typifies the possible future of plagiarism detection. "Reiterative" data analysis included data collected from *interviews*, *documents analysis*, and *screen casts*. Unpublished-data of interest also included *eye-tracking*.

Iterative data analysis refers to a systematic, repetitive, and recursive process of analysing results through clustering. Clustering involves putting similar items of interests together into groups or classes (McNabb, 2015, p. 290). Iterative sampling, on the other hand, ensures that the information-rich participants are included appropriately in the study. Reiterative data analysis involves reviewing the data analysis, and sometimes revising its collection and use. This action is sometimes mistakenly criticized as "lacking rigor", a type of "cherry picking". However, the 're'-iterative approach actually results in providing improve sensitivity to Big Data, together with its richness and variability (Mills, Durepos, & Wiebe, 2009, p. 504).

The results of Big Data

The iterative data analysis of screen cast and eye-tracking data is of particular importance to monitoring ethical writing. The data from screen casts were transcribed by a team of research assistants and included time based records on:

- what actions were performed – opening web page, key logging, and mouse movements,
- what resources were accessed and what information was copied, transcribed, translated, re-worded, or rephrased, and
- the computational resources that were employed to perform the aforementioned tasks.

Big Data Discussion and Outcomes

It should be evident, albeit implicitly, that these data could be useful in providing essential information about the development of writing skills, indicating the quality of the sources of knowledge, and providing valuable information throughout the writing processes. Placing importance on assessing the writer based on the measurement of these variables would be effective in de-emphasizing the effect of plagiarism. The desire to finish and submit a paper produced outside of the assessment measurement network would therefore be removed, or at-least, significantly reduced.

The transcriptions could be used to determine language processing times, the effective and fluent use of technology assisted writing applications, the use of unethical writing assistance, the dependency on translation and its use and subsequent incorporation into the text. The interactive analysis was then compared with document data to measure plagiarism management skills. Unfortunately, for this teacher, the student used the eraser when confronted with the challenge of correction. However, as a source of formative feedback, these results indicate the requirements for further training an in what area of the writing process it should be administered.

Although not mentioned in the results, eye-tracking data revealed results that have yet to be studied with any rigor. The eye tracking equipment was simply the camera, which comes standard in any computer. The camera data were only trialed and used with a few participants. Eye positional information and eye movement, that is tracking were the data of interest. Through inferential data analysis of the eye-tracking data with iterative data analysis of screen-cast transcriptions, the researcher claimed that the intent to cheat could be inferred. If this is in fact reliable, it would provide yet another valuable weapon in the war on plagiarism. If this hypothesis is correct, the intention to cheat could be used to trigger scaffolding devices to assist students. The assistance could be in the form of automated messages, or pre-emptive corrective feedback, or provide the tools to reduce the burden presented at that stage in writing.

The qualitative data from interviews indicated that even though the students were aware that their computers were recording their every move, some thought they would not be caught, particularly if they used quotations. In fact, some students in this study thought quotations were the easiest way to avoid plagiarism (Vasilopoulos, 2018). Other students resorted to using a paper dictionary, to avoid detection, while others

used Google Scholar and other search engines to find citations in [what they thought was] "relevant literature". These acts of plagiarism are undetectable using present technology. Therefore, the adoption of these, and other, covert means of monitoring student learning serve as a further illustration of how iterative data analysis can be effective in winning the war on plagiarism, for the time being at least.

Conclusion

This use of Big Data should now be recognisable as effective in de-emphasizing plagiarism. This data could also be summarised and reduced to provide verification of originality with credibility beyond that offered by present-day originality reports.

Covert data collection on the web

This final technique brings us closer to the 'ideal world' that was discussed in Todd (2010). If students were convinced that everything they do would be seen and reported prior to submitting their work, then possibly what presently constitutes cheating and dishonesty would become impossible. This seems contradictory to the findings of Vasilopoulos, because some students still thought they would not get caught. However, if the quality of the sources, and the process were to play an important role in assessment, and if feedback were provided progressively, copying would become an indication of incomplete training, and not an act of dishonesty (Todd, 2010, p. 148), and the process then becomes the important factor in assessing the quality of a finished product (Barrie, 2008).

The only weakness in the techniques so far is that the technology is not sensitive to a counterpart doing the work on your device. This is where the ethics of Big Data use becomes confrontational. The solution presented in this section verges into the realm of invasion of privacy, and extends beyond the use of school computers, intranet, and library searches. The data of interest in this section looks at the communications between people, its severity, frequency, and context.

Introduction to AI and NLP

The logical progression from monitoring individuals and their devices is to look at their interactions, the content of their communications, and their reactions with peers through various forms of communications. One such study (Kramer & Russell, 2014) analysed the unstructured data from Emails with Natural Language Processing (NLP) and Dynamic Graph Analysis. Mentioned earlier, dynamic graph analysis includes the clustering of similar data of interest.

NLP has been used in language studies, particularly useful is using NLP to conduct pragmatic analysis of the study structures of sets of text in an effort to understand the actual meaning. For computers, NLP is the closest thing there is to enabling computers with verbal and listener qualities. Mostly, NLP research has centred around been on the semantic structures of language. However, pragmatics are proving to be one of the greatest areas of research interest. One reason is possibly that NLP suffers also from the problem of understanding "verbal behaviour" (c.f. Skinner, 1957, as cited in

Cherpas, 1992). The resultant confounds that pragmatics and verbal behaviour place on AI using NLP indicate the importance of programming the Artificial Intelligence with a 'frequent user'. This means, high frequency consistencies in user communication profiles are necessary to program the AI to learn.

Another weakness in the AI system is that NLP can be fooled using morph-syntax, semantics, pragmatics and world knowledge (Mahler, et al., 2017). In fact, the 'fooling' is a miss-programming of algorithm sentiment analysis systems. This affects the trustworthiness of the data, its validity and reliability. However, this is not all bad news. If writers are required to use dedicated machines and internet service providers (ISPs), the AI can in fact become sensitive to the actual author. Therefore, NLP would be able to detect a third party on the system and produce an authenticity report to verify the authenticity of authorship.

Method

Researchers (Kramer & Russell, 2014) provide an in-depth background into the effectiveness of NLP as a language learning analytical tool. However, for the purpose of this paper, the area of interest surrounds the data analysis of business correspondence via emails, exactly 343, 134 in fact. The language processing artefacts included messages, sentences, phrases, tokens, co-references and their chains. The data in fact is large. It constitutes Big Data analysis and therefore calls for advanced analytics, as mentioned previously.

Results

By studying the semantic connections of Email addresses, NLP can identify individuals that are most influential and able to disseminate information using a K-Core Decomposition Model (Kramer & Russell, 2014, slide-22). Word level analysis from within text is also possible. Kramer and Russel provided evidence of the time, place, and communicator of phrases with positive emotional content, and phrases that showed anxiety. From these two sets of different kinds of data, the addresses and types of emotions can be mapped over time to see the flow of emotions, through a network.

Discussion

With this level of analysis available, it seems irrational that we cannot source the flow of text-based information in a writers' research networks. This network includes not only the searches and communications with data bases and library resources, but also through emails, social media, and now, virtual "face-to-face" meetings online. The advantage of NLP is that it can be done synchronously, or forensically. In fact, Google AI uses semantic analysis to predict writing, and search patterns. The implications of this is that Google AI could influence the untrained researcher into areas where Google deems important. Considering the number of untrained researchers far outweighs the number of critical researchers, we also need to be wary of this influence on academic integrity in the future.

Conclusion

NLP has the capacity to track and report on the flow of information either synchronously or forensically. The knowledge of this affordance should strike fear into the hearts of those who prey on desperate students, and provide services that rob them of valuable writing experiences.

Trolls, tricksters, and threats

Concerns about contract cheating date-back to the early days of distance education via school of the air in remote areas of Australia in the early 1950s. The saving grace in those times was the remoteness and delivery method was credible, through the Royal Flying Doctor service. Later, lessons were conducted for about one hour per day via short wave radio. Therefore, the only 'contract cheating' would be from siblings or relatives. The other saving grace was the pedagogical approach, children would be given tasks, that were non-repetitive from year to year. The earliest reported cases of concerns about contract cheating date back to the early 1950s, where "contract cheating was observed on [a US] campus" (Lancaster & Clarke, 2016, as cited in Nair & Saiz-Álvarez, 2019, p. 12).

This section introduces the opposition in the battle on academic integrity, the tricksters who crouch near the door of those who are credible, often in an effort to take on their appearance. Lewis Hyde (1998) writes that tricksters hide around the doorways of credibility because it is "one of the places where deep-change accidents occur" (p. 124). One place where "deep change accidents occur are in writing centres" (Geller, Eodice, Condon, Carroll, & Boquet, 2007), and whether we are editors, reviewers, administrators, or writing teachers, we too are the gate-keepers, which is why we come into close proximity with these individuals, and services.

Tricksters, and the more sinister "Trolls" often use plagiarism detection, and originality devices to 'bait' students and writers into using their services. Normally, teachers and reviewers provide the necessary feedback to scaffold the writer into finding a solution to writing problems. This exploration is helpful in learning, making ethical decisions, and other writing skill development. The services that "Trolls" provide are usually of a standard beyond the writer's peers (Lines, 2016). However, in crossing the boundary from learning to merely completing (Hyde, 1998), the trolls are effectively robbing the learners from experiencing these crucial developmental stages in writing. As a result, the trolls reinforce a dependency and this in turn forms unhealthy habits, and the need to return for services in the future.

A study into essay mills (Clarke & Lancaster, 2006) investigated the repeated use of a service provided by RentACoder. The study of more than 200 users found that only 8% of clients requested work in a one-off contract. A significant proportion (64.8%) of the clients ($n = 153$) made requests from RentACoder at least twice and up to 10 times. One hypothesis for the single uses was that new users are entering all the time. However, the discussion presented in this paper indicates that there are developmental and pedagogical reasons why writers resort to copying, or cheating. My hypothesis is, that the 8% of one-off students comprise predominantly of individuals who were desperate at the time and did not want to jeopardise their grades. Later, these individuals possibly felt guilty or ashamed and made concerted efforts to later developed the necessary

skills and language to avoid the same transgression in the future; a type of negative reinforcement or motivation. The rationale behind this belief is that most students eventually want to peruse a life that is ethical and morally sound. So, having to justify dishonesty in the future would certainly disrupt their future aspirations.

Another problem is that education and the trolls have shifting boundaries on what defines plagiarism. An "aggregated" definition identifies plagiarism as an action that hurts. It can be identified as *the use of another person's work* (USQ, 2017) *and passing it off as your own* (uOttawa, 2019), *which may result in a false academic evaluation* (uOttawa, 2017). "The trolls, the contract cheating services and the essay mills prepare work for students, who in turn submit the work as their own. This endeavour in turn benefits" all of those involved [directly] in the process (Lancaster, 2016), but also leads to a false evaluation of honest students who have submitted less superior material. Above and beyond this false evaluation of innocent students, contract cheating represents a serious threat to academic integrity (Lancaster & Clarke, 2017) on the grounds of ethics, morals and justice.

The social transition that writers experience when they transgress jeopardises their legal privileges and places them at risk of academic penalties, withdrawal of published material, and expulsion. Many of these have been widely publicise, but not too many are familiar with the legal implications. One example case (A.V. v. iParadigms LLC, as cited in Todd, 2010, p. 143) was where students claimed that the archiving and use of archived information was in breach of copyright. The case was held up initially, but an appeal submitted by iParadigms, the produces Turnitin®, was honoured on grounds of "fair use".

However, the boundaries of "fair use" cannot be assumed equal (Todd, 2010). The trolls, essay-mills, ghost-writer and other contract cheating services use the data collected during plagiarism checking together with work from contracted writers to deceive and provide personal gain for their clients. This action effectively removes legal rights from the individual who crosses this boundary, because no venerable court, in the UK at least, would offer assistance to an individual whose cause of action was based upon an "immoral or illegal act" (Holman v. Johnson, 1775, as cited in Todd, 2010, pp. 143–144.

Conclusion

Employing the use of writing services hurts students because it robs the perpetrates of a valuable developmental experience, forces them to transgress and become vulnerable and dependent on further malpractice, and robs innocent individuals of fair assessment.

Summary and Discussions

For more than a few decades, we have been struggling with the increased insurgence of contract cheating (Clarke & Lancaster, 2006). Merely identifying writing that was written produced by a writing contractor, or some online spin-writer, and punishing the writer should be considered unethical. The focus should be on where the system

has let the student down. This paper has presented empirical evidence to support the belief that we can win the war on plagiarism.

It must be remembered that students and some professional writers are aware that reviewers and markers can only check as thoroughly as time and other resources allow. However, this article has provided sufficient evidence to show this belief should no longer be allowed to exist. If technology is integrated with the writing process, it is possible to measure learning gain, processing of information, fluency in writing, and a host of other authoring metrics.

In summary, the present study found that by integrating technology and pedagogy and using formative assessment, writing tasks can be effective in reducing the requirements that make plagiarism an issue (Todd, 2010). In the second study, Vasilopoulos found that even when students knew they were being monitored, they still believed they could avoid plagiarism, simply by using quotations. The technological outcome of the Vasilopoulos study indicated that academic integrity could be automated through the covert collection of data, and it was proposed that this could result in a certified verification of authenticity based on the writing process and where information was sourced and how it was processed.

In the third study, the covert analysis of communication data was used as an example to illustrate how data flow through networks can be measured to indicate research integrity. The system can be used during the writing process or as a forensic device to investigate the validity of the process of writing over time.

Now that we are aware of these technological and pedagogical techniques, it would send a mixed message to the businesses, government, and our students who we both support and depend on for our mutual benefit. In fact, knowing now, the possibilities that are set before us, continuing on the present course would be unethical, because it would be providing a service that we are aware of that is inferior, while we know how it can be improved.

The present paper may be instrumental in winning the battle for now, however, the successor to plagiarism and contract cheating will possibly be cloud based and include Big Data analytics and AI. Therefore, perusing the development of these areas and further research in applying these services is definitely a priority.

Conclusions

Until now there has been a technological cold-war between the academics and the services that entrap writers to submit plagiarised material. This paper presented four progressive areas that are instrumental to the war on academic integrity. Without a doubt, the most influential weapon at our disposal remains in pedagogy, the expectations we place on our students, how we teach them, and how we assess learning. This paper suggests that the process, and the openness of sharing knowledge construction is paramount, which therefore, means that explicit understanding of the importance of integrity and collaboration needs to take a more prominent role in education. Three truly innovative approaches were also featured that should become the standard for licensed writing.

Platforms for writers should include internet activity which is openly shared with the author and the reviewers; without it, we are left to trust good referencing which, until now can only be checked for availability. There is no verifiable method, other than those presented in this article, to ensure the author actually understands the field or the research they have reported. We are left with language analysis, honesty, referencing standards, and search engines to depend upon. These are merely technical augmentations of age-old approaches.

The resounding message throughout this paper surrounds our ethical duty to maintain academic integrity. Now that this paper has offered a glimpse into the future and how we might be able to win this battle, and all the while reinforce and maintain the integrity of academic knowledge and education, it would send a mixed message to the world (i.e. businesses, government, and our students) who both support and depend on us, if we refrained for a moment on taking necessary action.

References

Barrie, J. M. (2008). How original: Catching the cheats. *The Biochemist, 30*(6), 16–19. Retrieved from https://portlandpress.com/biochemist/article-pdf/30/6/16/3906/bio030060016.pdf

Bretag, T. (2020). Keynote Speech. *Presented at Plagiarism across Europe and Beyond, 6th International Conference*, April 17–19, University of Wollongong, Dubai.

Cherpas, C. (1992). Natural language processing, pragmatics, and verbal behavior. *The Analysis of verbal behavior, 10*(1), 135–147. doi:10.1007/BF03392880

Clarke, R., & Lancaster, T. (2006, June). Eliminating the successor to plagiarism? Identifying the usage of contract cheating sites. In *Proceedings of 2nd international plagiarism conference* (pp. 1–13). Northumbria Learning Press. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.5440&rep=rep1&type=pdf

Dusza, D. G. (2019, October). Preventing Plagiarism – An Integrated Technology, Pedagogy, and Content Knowledge Approach [Presentation]. Presented at *The 27th Annual Korea TESOL International Conference*, Sookmyung Women's University, Seoul, Korea.

Dusza, D. G. (2020, January). Preventing Plagiarism with Integrated Technology, Pedagogy, and Content Knowledge. In *Proceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning* (pp. 116–120). doi:10.1145/3377571.3377576

Geller, A. E., Eodice, M., Condon, F., Carroll, M., & Boquet, E. (2007). *Everyday writing center: A community of practice*. University Press of Colorado.

Hyde, L. (1998). *Trickster Makes This World: Mischief, Myth and Art*. New York: Farrar, Straus and Giroux.

Jones, M., & Sheridan, L. (2017). Back translation: an emerging sophisticated cyber strategy to subvert advances in 'digital age' plagiarism detection and prevention. *Assessment & Evaluation in Higher Education, 40*(5), 712–724. doi:10.1080/02602938.2014.950553

Kramer, S., & Russell, M. (2014). Got Chaos – Extracting Business Intelligence from Email with Natural Language Processing and Dynamic Graph Analysis [Online]. Retrieved from https://www.slideshare.net/dreasoning/digital-reasoning-paragon-science-data-daytexas-2014finalsmall

Lancaster, T. (2016). Presentation slides: Contract cheating and the essay writing industry – where does the money go? *ICAI Inaugural Mediterranean Conference Integrity as a Way Forward*. Athens, Greece. September 2016. https://www.slideshare.net/ThomasLancaster/contract-cheating-and-the-essay-writing-industry-where-does-the-money-go-international-center-for-academic-integrity-inaugural-mediterannean-conference-athens-greece-8-september-2016

Lancaster, T., & Clarke, R. (2017). Rethinking assessment by examination in the age of contract cheating. Published in *Plagiarism across Europe and Beyond 2017 – Conference Proceedings*, pp. 215–228. Retrieved from
http://academicintegrity.eu/conference/wp-content/uploads/2019/07/proceedings17.pdf

Lines, L. (2016). Ghostwriters guaranteeing grades? The quality of online ghostwriting services available to tertiary students in Australia. *Teaching in Higher Education, 21*(8), 889–914. doi:10.1080/13562517.2016.1198759

Mahler, T., Cheung, W., Elsner, M., King, D., de Marneffe, M. C., Shain, C., & White, M. (2017, September). Breaking nlp: Using morphosyntax, semantics, pragmatics and world knowledge to fool sentiment analysis systems. In *Proceedings of the First Workshop on Building Linguistically Generalizable NLP Systems* (pp. 33–39). retrieved from
https://www.aclweb.org/anthology/W17-5405.pdf

McNabb, D. E. (2015). *Research methods for political science: Quantitative and qualitative methods*. Routledge. doi:10.4324/9781315701141

Mills, A. J., Durepos, G., & Wiebe, E. (Eds.). (2009). *Encyclopedia of case study research*. Sage Publications. doi:10.4135/9781412957397.n185

Nair, S. R., & Saiz-Álvarez, J. (2019). *Handbook of Research on Ethics, Entrepreneurship, and Governance in Higher Education*. IGI Global. http://doi:10.4018/978-1-5225-5837-8

Rogerson, A., & McCarthy, G. (2017). Using internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism? *International Journal of Educational Integrity, 13*(2), 1–15. doi:10.1007/s40979-016-0013-y

Skaar, H. (2015). Writing and pseudo-writing from Internet-based sources: implications for learning and assessment. *Literacy, 49*(2), 69. doi:10.1111/lit.12045

Stamatatos, E. (2011). Plagiarism detection using stopword ngrams. *Journal of the American Society for Information Science and Technology, 62*(12), 2512–2527. doi:10.1002/asi.21630

Stapleton, P. (2010). Writing in an electronic age: A case study of L2 composing processes. *Journal of English for Academic Purposes, 9*(4), 295–307. Retrieved from
https://www.learntechlib.org/p/75621/

Todd, P. (2010). Plagiarism detection software: Legal and pedagogical issues. *The Law Teacher, 44*(2), 137–148. doi:10.1080/03069400.2010.486165

uOttawa. (2017). Academic Regulation 14 – Regulation on Academic Fraud: Definition [Online Article]. *Administration and Governance, University of Ottawa*. Retrieved from https://www.uottawa.ca/administration-and-governance/academic-regulation-14-other-important-information

uOttawa. (2019). Academic fraud [online article]. *Office of the Provost and Vice-President, Academic Affairs, University of Ottawa*. Retrieved form https://www.uottawa.ca/vice-president-academic/academic-regulations-explained/academic-fraud

USQ. (2017). Plagiarism [online article]. *University of Southern Queensland*. Retrieved from
https://www.usq.edu.au/library/referencing/plagiarism

Vaitsis, C., Hervatis, V., & Zary, N. (2016). Introduction to big data in education and its contribution to the quality improvement processes. In Soto, S. V., Luna, J., & Cano, A. (Eds.), *Big Data on Real-World Applications*, pp. 41–63. doi:10.5772/63896

Vasilopoulos, E. G. (2018). Writing to Avoid Plagiarism: Computer-Mediated Source-Based Writing in a Pre-University EAP Program. Presented at the *Symposium of Second Language Writing,* held at Simon Fraser University Vancouver, Canada, August 2–4.

Wrigley, S. (2019). Avoiding 'de-plagiarism': Exploring the affordances of handwriting in the essay-writing process. *Active Learning in Higher Education, 20*(2), 167–179. doi:10.1177/1469787417735611

WROBEL, M. (2019). Ethical Issues Related to Machine Translation in the University Setting. In D. Dusza, A review of selected presentations from *The 11th Symposium on Writing Centers in Asia*, with implications for Kanda Institute of Foreign Languages [unpublished article].

Author

Daniel Dusza, Kanda Institute of Foreign Languages, 3-12-17-301, 256-0816, Odawara, Japan, e-mail: `dusza-d@kifl.ac.jp`