Detection of Online Contract Cheating Through Stylometry: A Pilot Study

David C. Ison
Northcentral University, San Diego, CA

Abstract
Contract cheating, instances in which a student enlists someone other than themselves to produce coursework, has been identified as a growing problem within academic integrity literature and in news headlines. The percentage of students who have used this type of cheating has been reported to range between 6% and 15.7%. Generational sentiments about cheating and the prevalent accessibility of contract cheating providers online seems to only have exacerbated the issue. The problem is that no simple means have been identified and verified to detect contract cheating because available plagiarism detection software is ineffective in these cases. One method commonly used for authorship authentication in nonacademic settings, *stylometry*, has been suggested as a potential means for detection. Stylometry uses various attributes of documents to determine if they were written by the same individual. This pilot assessed the utility of three easy-to-use and readily available stylometry software systems to detect simulated cases of contract cheating academic documents. Average accuracy ranged from 33% to 88.9%. While more research is necessary, stylometry software appears to show significant promise for the potential detection of contract cheating.

Keywords: academic integrity, contract cheating, plagiarism, stylometry

Ison, D.C. (2020). Detection of online contract cheating through stylometry: A pilot study. *Online Learning, 24*(2), 142-165. https://doi.org/10.24059/olj.v24i2.2096

Detection of Online Contract Cheating Through Stylometry: A Pilot Study

Various forms of cheating have plagued education since the inception of assessments of learning. Higher education, in particular, has endeavored to maintain academic integrity to provide meaning to the degrees they confer. Online programs have struggled with student authentication in various ways yet there are limited means of verifying authorship beyond the use of plagiarism detection technologies (Newton, 2018; Singh & Remenyi, 2016; White, 2016). Higher education faces an additional hurdle in that student participation in courses is remote and it cannot always be assured that the registered student is the individual logging into the course and completing assignments. In attempts to minimize cheating on traditional assessments such as quizzes and exams, higher education institutions often shift more focus on written assignments as they require

a higher level of thought and reasoning, and, in theory, are harder to thwart through simplistic cheating like copying from neighbors (Bretag et al., 2018; Newton, 2018; Singh & Remenyi, 2016; White, 2016). Even written assignments have been subject to various forms of duplicitous student activities. *Plagiarism*, the use of the words of another rather than the author's words or ideas, has been prevalent through the history of higher education. However, with the increased use of sophisticated plagiarism detection, the cutting and pasting of material from sources has become an unreliable and risky alternative for students. *Contract cheating*, also referred to as ghostwriting, is when a student employs an individual to complete assignments for them, has subsequently become more appealing. The benefit of using a contractor to complete coursework or writing assignments for students is that the resulting document is generally original thus not subject to detection by text-matching software (Lancaster, 2019, Lines, 2106). Current data indicates that "up to 16 percent of students have paid someone to do their work and that the number is rising" (Smith, 2019, para. 17).

Contract cheating is considered a relatively new form of cheating, estimated to only be a few hundred years old. While sharing of previous assignments and enlisting the help of others to complete work has been around for some time, such as within the U.S. fraternities and sororities, the 1940s seemed to mark a turning point for contract cheating. At this time, there was a significant uptick in the numbers of advertisements for ghostwriting services in New York City newspapers. Proliferation continued through the 1960s and 1970s, which included solicitations for assistance in writing papers, theses, and dissertations (White, 2016). While these services became more widely available, there was still the problem of the student finding the contractor as well as completing payment. This dynamic completely changed with the advent of the Internet. Services became easily discoverable and payment could be made anonymously to anyone around the globe. As such, essay or paper mills, in addition to individual writers, became a realistic and timely option for students looking to circumvent work (Singh & Remenyi, 2016; White, 2016).

According to Lines (2016), the use of such ghostwriting services has been steadily increasing over time:

> while collusion in the form of paying another person to complete all or part of an assessment is not a new form of cheating, technological advances and changes in the socio-economic context of tertiary education in recent years have led to a worrying trend in which this practice appears to be becoming more widespread. (p. 889–890)

Lancaster (2019) estimated the 2014 revenue for the industry to be in excess of $100 million. Both supply and demand appear to be very healthy for this form of commerce.

In a study by Rigby, Burton, Balcombe, Bateman, and Mulatu (2015), students were very willing to purchase papers. The emphasis on high grades, as well as the economic incentive of advanced degrees (which can significantly increase lifetime compensation in the workplace) is a possible explanation (Lines, 2016; Rigby et al., 2015). While the one-time or occasional use of ghostwriting for written course assignments could be problematic but not cataclysmic, the notion of a student receiving credit for something they did not do is disturbing. More importantly, though, is the award of degrees in which students have bypassed necessary learning through employing the services of another person (Lancaster, 2019).

As graduate school provides critical research education and skills that provide the basis for careers in science and academia, circumventing this learning could foster erroneous research

practices (Singh & Remenyi, 2016). For example, it is now possible for someone to have their dissertation written, in its entirety, for $3,000 to $5,000 (Top 20 Writing Services, 2019). An individual can be awarded one of the most coveted degrees in the world, the doctorate, with little or no original work by the student (Lancaster, 2019; Lines, 2016). The issue of contract cheating is so rampant that PayPal has recently refused to pay known contract cheating providers through their service, though after six weeks of the embargo many essay-writing services are still successfully accepting PayPal payments (Bailey, 2019; Coughlan, 2019).

Even in light of the apparent rise in contract cheating, little empirical evidence exists on its prevalence or impact within higher education. Unfortunately, an accurate estimate is difficult to garner as data on the actual prevalence of this type of cheating only comes to light if or when a student is caught (Bretag et al., 2018; Harper et al., 2018; Lines, 2016; Singh & Remenyi, 2016). Available data does, however, indicate that contract cheating is a concerning issue in the postsecondary sector. Although the incidence of students partaking in contract cheating is still small (reported to be approximately 3% in the U.S.), this estimate means that 1.7 million students are purchasing papers or other services (National Center for Education Statistics, 2019; Wallace & Newton, 2014).

The one consensus about contract cheating among researchers is that it is challenging to detect. Current text-matching detection applications are ineffective in identifying this type of cheating (Anekwe, 2010; Dawson & Sutherland-Smith, 2018; 2019; Lancaster, 2019; Lines, 2016; Newton, 2018; Rogerson, 2014; 2017; Singh & Remenyi, 2016). While Turnitin has recently launched its proprietary Authorship Investigate service that aims to detect contract cheating, it has yet to be widely adopted by higher education institutions (Turnitin, 2019). As of yet, this new Turnitin service has not been vetted through empirical scholarly research (Singleton & Ricksen, 2019). Moreover, it appears the service relies on Turnitin's database of previous student work, which may not include all assignments or that of newer students (Turnitin, 2019).

As Rogerson (2017) stated, the lack of effort on the part of higher education stakeholders to act against contract cheating is problematic. The claim that contract cheating is difficult, if not impossible, to detect has not been explored by exigent research and very little empirical evidence of the effectiveness of contract cheating detection methods exists. Because of these facts, Singh and Remenyi (2016) specifically call for a proactive intervention, as

> academic cheating undermines the good name of the institution and calls into question the integrity of both the faculty and students. There is every reason for a university to take all forms of cheating seriously, and to eliminate it wherever possible. (p. 36)

Thus with the increasing threat of contract cheating, it is time for higher education to address how to detect and handle this category of academic integrity violation (Medway, Roper, & Gullooly, 2018; Newton, 2018; Slade, Rowland, & McGrath, 2019).

One promising tool that can be used in detecting ghostwritten work is *stylometry*, the study of the writing characteristics of a specific author. To address contract cheating, Juola (2017) noted that "stylometry is an important and relatively mature technology that can be usefully applied to address a key problem in education" (p. 196). Dawson and Sutherland-Smith (2019) reiterated that research is needed in order to "focus on approaches to improve detection rates" (p. 291) specifically through empirical studies. Higher education needs "a process approach… to identify, document, and investigate irregularities using technological, interpretive, and conversational

means" (Rogerson, 2017, p. 3). Therefore, this study set out to explore the utility of user-friendly stylometry analysis software for the detection of contract cheating of written assignments. The following research questions were used to guide this study:

RQ1: What is the accuracy of user-friendly stylometry analysis software?

RQ2: Which of these software performed with the highest accuracy?

Review of Relevant Literature

Defining Contract Cheating

Singh and Remenyi (2016) defined contract cheating as the "practice of hiring a writer (or writers) to produce a piece of work that follows a predefined style, and none of the original writing credit is attributed to the ghostwriter" (p. 37). Therefore, contract cheating is not generally classified as plagiarism, in terms of stealing or theft of words or ideas. Instead, it is authorship fraud, stating that the identified author wrote the work. Other researchers have used more straightforward definitions, stating that anytime a student outsources assignments to fall within tailored instructions of the student, they have conducted contract cheating (Bretag et al., 2018; Clare, Walker, & Hobson, 2017; Harper et al., 2018; Lancaster, 2019; Lines, 2016). While contract cheating and ghostwriting are often used synonymously, Rogerson (2017) used a more specific term, "cyber-pseudepigraphy" (p. 1), to describe these practices through the use of the Internet. While there are minor differences in interpretations as to what constitutes contract cheating, or ghostwriting, the consensus among researchers is that it is an egregious form of cheating (Bretag et al., 2018; Clare, Walker, & Hobson, 2017; Harper et al., 2018; Lancaster, 2019; Lines, 2016; Rogerson, 2017).

Contract Cheating, Ghostwriting, and Paper Mills

In certain writing genres, ghostwriting and the use of pseudonyms have been commonplace, such as in fiction and among autobiographies (Farhi, 2014). Rarely does this practice seem to be problematic for authors or readers when agreements concerning terms of ownership and compensation are handled transparently. Such practices are also used in medical writing with pharmaceutical companies hiring authors to write studies on their medications and honorarily listing prestigious doctors as the actual authors. According to the National Institutes of Health, anywhere from 10 to 40% of research articles involving pharmaceuticals may be ghostwritten. Such practices are generally frowned upon, and in some cases subject to violations of ethical agreements with funding agencies, yet they continue to occur (Anekwe, 2010). Much less tolerance seems to exist for the use of contract or ghost writers by students to complete academic requirements (Lines, 2016).

Prevalence and Trends in Contract Cheating. A limited number of studies have determined the prevalence and trends of contract cheating in academia. Newton (2018) reviewed 65 studies dating from 1978 through 2014 for student self-reported conduct of contract cheating. In 1978, the average prevalence among respondents was 3.52%, increasing steadily to 15.7% in 2014, which was estimated to equate to approximately 31 million students. Although the details about the studies and their potential equivalence for comparison were lacking, it does appear that contract cheating is increasing among students and the absolute number of potential participants is disconcerting. In 2013, Turnitin conducted a study in which 7% of student respondents admitted to purchasing at least one assignment. Rigby et al. (2015) found that under typical academic

circumstances, 50% of the sample of students reported they would be willing to risk purchasing a written assignment. Two recent studies were completed in Australia, one querying academic staff and the other focused on students. The survey of 916 Australian academic staff discovered that 66% of respondents believed they have had students submit ghostwritten work on at least one occasion and 40% believed this had occurred five or more times (Harper et al., 2018). Among the 814 Australian students surveyed, 6% admitted to partaking in contract cheating (Bretag et al., 2018). Discussion within the literature suggested that these percentages are likely an underestimate, as students may not truthfully report transgressions (Bretag et al., 2018; Harper et al., 2018; Lines, 2016; Ma, Wan, & Lu, 2008; Rigby et al., 2015; Wallace & Newton, 2014).

Detection of Contract Cheating

As the existing literature has noted, contract cheating is very difficult, if not impossible, to detect with precision with current methods and tools. In particular, the fact that ghostwritten works can be of relatively good quality, or at least sufficient enough to receive a passing grade, complicates matters. Because insufficient research supports a specific method of detection, Rogerson (2017) stated "there is a need for an evolutionary approach to enhance evaluation skills beyond discipline related practices and academic writing conventions…. An approach that can streamline methods of determining irregularities and documenting evidence for evaluation and discussion" (p. 2). The literature has advocated only limited approaches to exposing contract cheating. These include looking for clues of radical differences in student performance across assignments, use of assessment data for comparison of assignments within courses and institution-wide, as well as ensuring proper documentation of past violations of academic integrity as student cheaters were found to act in a nonrandom fashion and likely to be repeat offenders (Clare, Walker, & Hobson, 2017; Taylor, 2014; Rogerson, 2017). The use of software and computer aids has also been suggested to confront contract cheating. Throughout the literature, no consensus exists on the most effective means of detection; moreover, what methods do show promise may suggest potential ghostwriting but do not necessarily prove it (Lines, 2016; Rogerson, 2014; 2017).

Evaluator-Based Detection. Various researchers have recommended that observations of inconsistent student writing by evaluators (i.e., faculty or teaching assistants) provide a means for detecting contract cheating. Lines (2016) suggested that student writing styles need to be tracked, though this technique may be unable to distinguish between genuine student work (albeit it significantly improved or of poor quality) when compared to past performance. Also, as noted by Singh and Remenyi (2016), graders and even faculty may not be adequately familiar with the student's writing style and quality of work to notice. Further, universities do not appear to be directing enough resources to ensure that students are the ones completing coursework. Overworked faculty and teaching assistants also may not have the time or energy to commit to unassisted attempts to "hand" detect such occurrences (Singh & Remeny, 2016).

In a study of products of essay mills in the U.K., Medway, Roper, and Gillooly (2018) purchased two papers to be evaluated by ten evaluators. The participants were not informed of the goal of the study. Not one of the graders detected issues with either paper, although these individuals did not have reference to previous works of writers. Dawson and Sutherland-Smith (2018) completed a study which provided seven graders with 20 assignments to grade in two groups of ten assignments of which three were purchased by an online provider. Participants were explicitly asked to identify which papers were a product of contract cheating or written by a "real" student. The average accuracy of detection was 57.2%. If considering only the contract papers, markers were able to identify 62% of the works. In 11% of cases, markers wrongly identified a

student paper as ghostwritten. In a follow-up study, Dawson and Sutherland-Smith (2019) enlisted fifteen graders which were given 20 papers, six of which were purchased, to evaluate. The graders were then given training on identification of contract cheating and subsequently given an additional 20 papers to evaluate, again with six of those provided being purchased online. True positive (correct identification of contract paper) improved from 17.3% pretraining to 24.6% posttraining. False negative (failing to detect a contract paper) dropped from 12.6% to 5.3%. Overall, papers were correctly identified as either student or contract work 75.3% before training and 85.7% after training. The presentation of results was sometimes confusing as some values did not include all papers which inflated reported values (Dawson & Sutherland-Smith, 2019).

Rogerson (2017) found similar difficulties in detection by graders noting that humans may or may not perceive writing irregularities resulting in an inability to distinguish between contract and low-quality student writing. Further, inconsistencies or incongruities do not equate to any level of assurance of contract cheating. Even if a paper is suspected of being ghostwritten by a grader, it would be hard to prove or take any action (Rogerson, 2017). Dawson and Sutherland-Smith (2019) came to a similar conclusion stating that "marker detection alone is not necessarily sufficient evidence to satisfy the burden of proof for a contract cheating allegation" (p. 722).

Other techniques suggested as possible solutions for the detection of contract cheating can be likened to "clue sleuthing." Both Lines (2016) and Rogerson (2014) found that ghostwritten assignments often include vague answers and lack the requisite depth. Additionally, ghostwriters frequently do not follow assignment instructions (Rogerson, 2014). Document metadata may also reveal clues such as the revelation of the document author information not matching the name of the student (Lines, 2016). Dramatic improvements in writing or language can also raise caution flags. Outliers in proposed grades were deemed to be suspicious. Other language-based warning signs include use of spelling or colloquialisms not common in the native language of the student. For example, the use of British English versus that of the U.S. is hypothetically a signature of a ghostwriter (Lines, 2016; Rogerson, 2014; 2017). While text-matching detection software does not work well to directly catch ghostwritten work, researchers found that abnormally low Turnitin similarity indices are undoubtedly worthy of further investigation, particularly if the references section of the work are not flagged, assuming they have not been omitted from analysis (Lines 2016; Rogerson, 2017). An additional clue common among contracted submissions is inconsistencies in references. These may include odd, very outdated, incorrect, or made-up references. Also, it was found that ghostwriters sometimes mix various components of references (e.g., using the author and publication date from one article while using the title and journal details from another) in order to thwart attempts at plagiarism detection (Lines, 2016).

Computer-Assisted Detection. While available text-matching detection technologies have proven to be relatively effective at catching textual overlaps with references, no software product currently exists that specifically targets contract cheating. Within attempts to detect plagiarism through text-matching, a technological arms race has occurred in which education stakeholders rush to find solutions to cutting and pasting of reference material while students search for ways around such safeguards. Currently, available detection programs cannot adequately protect against ghostwriting (Dawson & Sutherland-Smith, 2018; Juola, 2017; Lines, 2016; Rogerson, 2017). Turnitin's Authorship Investigate is one of the first integrated systems to detect contract cheating but comes at an additional cost and has yet to face the scrutiny of researchers (Turnitin, 2019; Singleton & Ricksen, 2019). To date, it seems that universities are not willing to invest in the resources, human or financial, necessary to combat contract cheating

through available methods of detection. Moreover, faculty and their assistants often do not have the time or tools available to make such an undertaking reasonably timely (Juola, 2017).

Stylometry

One method of analysis that has potential for use in detection of contract cheating is *stylometry*, the analysis of authorial style and writing attributes. In theory, this method of inquiry can determine the differences between authors, thus if a difference exists between genuine student submissions and those suspected of being ghostwritten, education stakeholders potentially have actionable evidence (Juola, 2017).

Uses of Stylometry. Stylometry has been used in a range of applications since it was first introduced. Stylometry has been implemented in forensic authorship analysis, such as in criminal and civil lawsuits. Intelligence agencies have also adopted the technique to determine the authorship of threats, Internet activity, and other potentially malicious writings (Brocardo, Traore, & Woungang, 2015; Neal et al., 2018). Because people have unique authorial fingerprints or idiolects, one author can be distinguished from another, e.g., the use of "near" instead of "by" (Juola, 2017, p. 189). In the analysis of the Federalist Papers, it was found that Alexander Hamilton never used the word "whilst" and James Madison never used the word "while" (Juola, 2017, p. 192; Mosteller & Wallace, 1963). Stylometry techniques also successfully "outed" J. K. Rowling as the real author of works produced under a pseudonym (Juola, 2017).

The goals of stylometry can vary depending on the type of outcome required by a researcher. In authorial attribution, the aim is "to determine the probability that a document was written by a particular author based on stylistic traits rather than the content of the document" (Neal et al., 2018, p. 86). Authorial verification entails a "binary classification problem that decides if two documents were written by the same author" (Neal et al., 2018, p. 86). When an author attempts to disguise their identity, termed obfuscation or adversarial stylometry, they will attempt to mask authorial style (Neal et al., 2018).

Stylometry Techniques. Various techniques of textual analysis exist within stylometry. Lexical analysis incurs the use of word-based and character features, which has the advantage of being robust against text "noise" (i.e., spelling and grammar errors). Lexical analysis includes word n-grams (i.e., repetition of words "n" times in a document, where "n" is the number of occurrences), word frequencies, words per sentence, the number of sentences, and vocabulary richness. Structural features may also be considered such as indentations, misspellings, grammar, and words specific to particular social or cultural backgrounds (Neal et al., 2018; Sarwar, Li, Rakthanmanon, & Nutanong, 2018). Syntactic features, which include punctuation and parts of speech, can also be employed to strengthen analyses (Brocardo, Traore, & Woungang, 2015).

The availability of fast, powerful, and inexpensive computers has allowed for advanced techniques to be developed. Machine learning classifiers and clustering have been recently adopted in stylometry analysis. This technique uses an algorithm that mathematically describes proximity of different data-points allowing for the distinction between data that naturally should or should not be grouped together. Another contemporary method are neural networks, Chain Augmented Naïve Bayes (CAN), and nearest neighbor calculations. Neural networks enlist a committee of machines that vote on author identity. CAN combines Bayesian analysis and n-grams. CAN was used in the exploration of authorship of the Federalist Papers and using only three features, correctly identified the author with 95% accuracy. The premise of nearest neighbor techniques is to determine the intertextual distance from a sample of training documents. Used regularly in

authorship attribution, differences in word frequencies can be calculated through chi-square analysis or the presentation of *z*-scores (Neal et al., 2018).

Best practices in stylometry have also been developed through numerous studies. Sample documents should be preprocessed for normalization such as removing nonalphabetical characters, capitalization, citations, names, cities, and dates. Next, selected features should be extracted (e.g., parts of speech, n-grams). Classification is then completed via comparison to training document features. Finally, an output is provided that provides authorship probabilities or designations (Neal et al., 2018; Prasad, Narsimha, Reddy, & Babu, 2015). Neal et al. (2018) determined that the accuracy was maximal when numerous examples for each author existed and when the total number of potential authors was small. Findings showed that documents to be analyzed should be on similar topics or from the same genre (e.g., a journal article vs. a mystery novel) and share authorial sentiment.

Athira and Thampi (2018) concluded that the ideal length of sample texts should be 500 words and when using this size of text, Brocardo, Traore, and Woungang (2015) had a 95.7% accuracy rate. A representative scenario for evaluation of authorship was noted to be composed of a training sample of 20 text blocks written by the known author vis á vis five blocks by the unknown author. Among the numerous types of analytics used by researchers, Sarwar, Li, Rakthanmanon, and Nutanong (2018) found that logistic regression with five documents per author provided sufficient accuracy for identifying authorship. In particular, the use of n-grams of the length of two to five with logistic regression was found to successfully delineate between authors (Prasad, Narsimha, Reddy, & Babu, 2015). Juola (2017) stated that a mixture of methods and feature extraction provides the most accurate results. Finally, it has also been established that simpler techniques are ideal, as these allow for more mainstream adoption and, in most cases, had a negligible impact on accuracy (Neal et al., 2018; Sarwar, Li, Rakthanmanon, & Nutanong, 2018).

Method

The purpose of this quantitative and descriptive pilot study was to evaluate commercial-off-the-shelf (COTS) stylometry software for its utility in potentially detecting contract cheating (Juola, Sofko, & Brennan, 2006; Salkind, 2012). The reasoning behind the use of COTS solutions is that they are readily available and relatively easy to use so that a student evaluator could employ them without a significant investment of time. Only free COTS software was included in this study to demonstrate that stakeholders would not be limited by financial investment restrictions (Juola, Sofko, & Brennan, 2006). An additional goal of this study was to assess stylometry software on a scenario that closely replicates how a student may submit work, both legitimate and ghostwritten.

Sampling Procedures

In order to simulate a real-world scenario in which an individual attempts to pass contracted work as their own, two corpora were developed. The first corpus comprised a random selection of five text blocks of approximately 500 words each from five separate peer-reviewed journal articles genuinely written by a known author.

The second corpus was created through the collection of five randomly selected text blocks of approximately 500 words each from five separate peer-reviewed articles from the same journal written by individuals other than the "known" author. All corpora were extracted from the same journal, the *Journal of Aviation Technology and Engineering*, which is highly focused on a specific

discipline and is well-respected among scholars in the subject area. Further, the selection of articles from one focused journal was conducted to retain both context and topic areas per the recommendations of Neal et al. (2018). This second corpus was labeled as "other" author or someone other than the "known."

Block lengths (500 words) were chosen as these were endorsed by exigent research (Athira & Thampi, 2018; Brocardo, Traore, & Woungang, 2015; García & Martín, 2012; Puig, Font, & Ginebra, 2016). The text blocks from both corpora were collected in a way to retain the last complete sentence to allow as close to 500 words as possible to ensure that an incomplete sentence did not influence results (García & Martín, 2012). The number of source original documents from which text blocks were selected exceeded the minimum document and text block counts recommended by Prasad, Narsimha, Reddy, and Babu, (2015). This allowed for additional text blocks to be available as test documents.

Test documents were randomly selected from both corpora to determine the ability of software to detect authorial differences. When a text block was used from a specific document, all other text blocks from that document were excluded from the known or other corpus during the specific analysis. Various combinations of corpora and test documents were examined. The most realistic contract cheating scenario would likely be the case in which only a single genuine work by a student (for example, something written in a controlled environment) would be available as a known author corpus. To test the software in this type of scenario, four text blocks (approximately 2,000 words) from a known author document were used to compare to four text blocks from a document written by a different author. All text blocks written by authors not included in the comparison were lumped into the "other" group for the test run (Juola, 2017; Prasad, Narsimha, Reddy, & Babu, 2015).

Measures

Three COTS software packages were used to evaluate text blocks. These were chosen specifically because each are backed by research implying positive results in author attribution as well as being free and readily available to download from the Internet. Moreover, with some brief reading of the programs' instructions, their use is intuitive even for persons without backgrounds in computer languages or coding. The first software was *Signature Stylometry System 1.0* (SSS) (Dawes, Merivale, & Millican, 2003). This system was developed to evaluate the Federalist Papers for authorship. Although this software is the most limited in capabilities, it is very simple to use and requires minimal computer memory processing power (Nieto, Sierra, Juan, Barco, & Cueto, 2008). The statistical analysis results provided by SSS are various chi-square comparisons of two documents or two corpora based on word lengths, sentence lengths, paragraph lengths, letter counts, and punctuation counts. Also, SSS provided graphical depictions of these measures (Dawes, Merivale, & Millican, 2003; Millican, 2003).

The second software, *Java Graphical Authorship Attribution Program* (JGAAP), was developed by researchers at Duquesne University and has been suggested as a potential means of detecting contract cheating (EVL Labs, 2018, Juola, 2017; Juola, Sofko, & Brennan, 2006). This program has a significant number of options available for analysis. Text blocks are loaded into the program from which point document analysis options are chosen: canonicizers, event drivers, event culling, and analysis methods. Canonicizers standardize the format of text such as normalizing whitespace, removing numbers, or making text all lowercase. Event drivers are the items the software seeks for the actual analysis, which can include parts of speech, n-grams, lexical

frequencies (a measure of reader processing of text), and first words of a sentence. Event culling allows to select how event drivers may be used (for example, looking only for the most frequent, a value set by the user, occurrences). The analysis method refers to the statistical processing of the resultant data. For this study, Naïve Bayes and K-Nearest Neighbors (KNN) were the selected methods of analysis. Within each of these categories, there is a wide range of additional options. One downside to JGAAP is beyond a certain level of analysis complexity, such as using more than a dozen event drivers, computer memory issues can become a factor, preventing the completion of analysis by a typical desktop or laptop computer (EVL Labs, 2018).

The third software used, *JStylo Authorship Attribution Framework v1.2* (JStylo), was created by the Privacy, Security, and Automation Lab (PSAL) at Drexel University and is an extension of JGAAP (Stolerman & Dutko, 2013). Much like JGAAP, JStylo has numerous options available for analysis. Documents are uploaded and users can choose from a range of features and classifiers to be examined during the analysis. Features include the items the software examines for statistical analysis such as average syllables in a word, average sentence length, and reading ease scores. Classifiers refer to the statistical testing options; in this case Naïve Bayes and Simple Logistic Regression were selected. Just as with JGAAP, advanced complexity testing (exceeding approximately a dozen features) can overwhelm computer memory capabilities, preventing the completion of analysis (Stolerman & Dutko, 2013).

Recommendations from research utilizing these software packages guided the selection of program settings used in this study (Athira & Thampi, 2018; Brocardo, Traore, & Woungang, 2015; García & Martín, 2012; Juola, 2017; Juola, Sofko, & Brennan, 2006; Prasad, Narsimha, Reddy, & Babu, 2015; Puig, Font, & Ginebra, 2016). The individual options chosen for each software are outlined in Tables 1, 2, and 3 for SSS, JGAAP, and JStylo, respectively.

Table 1

Selected Analysis Options for Signature Stylometry System 1.0

Category	**Selected Options**
Corpora	• Known author • Test (unknown)
Features*	• Word lengths • Sentence lengths • Paragraph lengths • Letters • Punctuation
Classifiers	• Chi-square
Analysis	Chi-square columns selected had minimum counts of 5; separated others into lower/higher columns.

* Features are standard and cannot be modified by the user.

Table 2

Selected Analysis Options for Java Graphical Authorship Attribution Program

Category	Selected Options

Corpora	• Training corpus
	  o Known author
	  o Other than known author
	• Test (unknown)

Event Drivers[a]	• Rare words count • Binned naming times (Complexity) • Sentence count • Average sentence length • Average syllables in word • Character n-grams (n = 4) • Word n-grams (n = 2)
Event Culling	• Most common events[b]
Analysis Methods	• Naïve Bayes • K-Nearest Neighbor with Cosine Distance[b]

[a] These were matched as closely as possible to JStylo available options.
[b] Settings were selected per the recommendations of software instructions (EVL Labs, 2018).

Table 3

Selected Analysis Options for JStylo Authorship Attribution Framework v1.2

Category	Selected Options

Corpora	• Training corpus
	  o Known author
	  o Other than known author
	• Test (unknown)

Features[*]	• Unique words count • Complexity • Sentence count • Average sentence length • Average syllables in word • Gunning-Fog Readability Index • Character space • Letter space • Flesch Reading East Score • Character n-grams (n = 4) • Word n-grams (n = 2)
Classifiers	• Naïve Bayes • Simple Logistic Regression
Analysis	Train on training corpus and classify unknown test documents

[*] Features other than n-grams are standard for the most basic analysis in JStylo. N-grams were added per the recommendations of the literature (Posadas-Durán et al., 2017; Ragel, Herath, & Senanayake, 2013).

Research Design

This quantitative, descriptive study sought to determine the accuracy with which stylometry analysis software can identify if a work was not written by a known author in a simulation of contract cheating scenarios. Accuracy of detection was calculated using the below formula from Prasad, Narsimha, Reddy, and Babu (2015):

$$Accuracy = \frac{\text{Number Correctly Identified}}{\text{Total Number of Items Analyzed}}$$

The results from each test run for individual software packages were recorded in Microsoft Excel for the calculation of descriptive statistics.

Procedure. All three software packages were evaluated for their ability to distinguish between a known author and someone other than the known author. Due to differences in the way the programs work, different procedures had to be used with SSS versus JGAAP and JStylo. Since SSS has very simplistic capabilities, designed only to compare two documents or groups of documents rather than more complex corpora, the known author documents were loaded as one group to be compared with others. The selection of documents and testing procedures are outlined in Figures 1 and 2 in the Appendix. A simulation of a realistic contract cheating scenario in which only one document was compared to another was conducted through random pairing of documents written by different authors. Another assessment of SSS was made comparing all known documents to all unknown documents.

In the testing of JGAAP and JStylo, one document from both the known and other corpora was randomly excluded for each analysis. The training corpus comprised the four remaining documents from both known and other groups. A range of one to five text blocks from the excluded "other" author document were used to compare with the training corpus. Also, one test run was completed to also include a randomly selected text block from the known author. Additional simulations of realistic contract cheating scenarios were conducted through the random pairing of documents written by different authors in which only one document (n = 4 text blocks) was compared to another (n = 4 text blocks). All remaining unknown author documents were classified as "other" during these test runs. The selection of documents and testing procedures are outlined in Figures 3 and 4 in the Appendix.

Data Analysis. In order to assess each software program, accuracy of results would need to be acquired. In each case, the software either correctly or incorrectly matched the author of the text being tested. The numbers of successful and unsuccessful times an author was identified were collected and summed. The process was repeated for each software package and method of statistical analysis. The aforementioned accuracy formula was then used to determine overall accuracy for each software.

Results

The results of the tests were mixed. JStylo performed well in all testing regimens while JGAAP presented inaccurate identification of known authors as the number of text blocks was reduced. SSS generally did poorly in all test types.

Statistics and Data Analysis

SSS Results. Due to the setup of SSS, accuracy was calculated as the number of significant chi-square tests versus the total number of tests for each round (n = 5). Overall, the test performance of SSS had an accuracy rate of 40%. During tests of individual documents, SSS performed with an accuracy between 20% to 40%, with an average of 33.3%. For the test comparing the entire known corpus with the unknown, SSS achieved an accuracy of 60%.

JGAAP Results. A total of 30 test runs were conducted using the test documents. Ten additional test runs were conducted using the four blocks of known and four blocks of unknown works. The overall accuracy for JGAAP when using Naïve Bayes analysis was 74.4% but only 43.7% when relying on K-Nearest Neighbor (KNN). Detailed analysis of the results of different numbers of blocks analyzed is presented in Table 4. During the testing to distinguish a known text block from an unknown block, JGAAP performed poorly, with a maximum accuracy of 25%. It was only after manipulating the settings of the program that this was increased to 87.5%. To achieve this increase in accuracy, discrete lexical frequencies were added as events, the standard deviation culler (which sought items with the largest standard deviation) was added, and Burrows Δ (which quantifies differences in event drivers of different texts) was used as the analysis method. These were added in attempts to improve accuracy per the recommendations of settings offered by the authors of the software (EVL Labs, 2018).

Table 4

Results for Tests Using JGAAP

Number of Test Text Blocks	Accuracy % (Naïve Bayes)	Accuracy % (K-Nearest)
1	82.9	57.5
2	86.1	75.0
3	57.8	34.4
4	77.5	26.4
5	70.6	25.0

JStylo Results. A total of 30 test runs were conducted using the test documents. Ten additional test runs were conducted using the four blocks of known and four blocks of unknown works. JStylo provided the most consistent and accurate results across all tests. The overall accuracy for JStylo when using Naïve Bayes analysis was 86.3% and 88.9% when utilizing Simple Logistic Regression. Detailed analysis of the results of different numbers of blocks analyzed is presented in Table 5. JStylo was able to dependably identify the majority of cases (83.3%) when a known author text block was added during testing.

Table 5

Results for Tests Using JStylo

Number of Test Text Blocks	Accuracy % (Naïve Bayes)	Accuracy % (Log. Reg.)
1	80.0	80.0
2	90.0	90.0
3	86.7	86.7
4	90.0	95.0
5	88.0	92.0

To further assess JStylo software, the data available for processing was limited. During this more rigorous testing, the number of features was reduced to five including the substitution of one feature that was not used in primary testing (unique words, complexity, character n-grams, word n-grams, lexical frequencies [added]). The number of known text blocks was reduced to two and the number of other blocks was reduced to four. Lastly, only one text block was used as the unknown. JStylo was still able to identify the block correctly at 91% accuracy (82% for Naïve Bayes, 100% for Simple Logistic Regression), albeit this procedure was only repeated five times with different test blocks.

Discussion

Based on the results of this pilot study, COTS stylometry analysis software has significant promise to evaluate document authorship. This, in theory, could easily be extended to student submissions. In this study, academic texts of a specific context and genre were used to assess the software, thus furthering the advocation of applying these methods to detect contract cheating. Considering that current methods and tools to detect contract cheating are tediously time-consuming (such as when using graders) and ineffective (in the case of text-matching detection software) the use of stylometry may provide a path for dealing with the growing issue of ghostwritten student work.

Clearly, stakeholders must have confidence in contract cheating assessment tools in order to use them for academic integrity enforcement. Accusing a student of contract cheating is a serious matter that could potentially affect their future at an institution. Further, it could potentially lead to legal action by the student to counter the claim. So even in light of the capacity of stylometry analysis, the variation in performance of the different software shows that careful selection of options and utilization are necessary to achieve reasonable accuracy.

It appears that SSS is not a good candidate to use in an academic setting. The software was easy to use and its interface was very intuitive, though text loading was more tedious than other software options. The very basic forms of analysis offered by the software are likely the reason why it performed so poorly. Moreover, the chi-square analysis did not work well due to many cells having inadequate counts to be an appropriate statistical tool even when combining cells. SSS does provide a graphical display of results, which can potentially be used to "eyeball" differences; however, stakeholders would likely balk at the idea of using this a primary means of detection. This feature could be a supplemental tool for assessors but not a primary means of identifying contract cheating. SSS seems to be more powerful when using much larger documents than used in this study though, again, this significantly limits its utility in a typical contract cheating scenario.

Although JGAAP is very sophisticated in terms of its background functioning, it was relatively easy to use. Uploading texts was efficient as was the selection of options. The challenge is to know which options to select, therefore the research literature was consulted for guidance. Further experimentation will likely yield more effective combinations of options. With an overall performance between 43.7% and 74.4%, JGAAP performed better than SSS. Naïve Bayes analysis, the stronger performer, would thus be the recommended setting for analyzing documents, though more thorough testing would be necessary. Although the performance of JGAAP in evaluating smaller quantities of text chunks was initially encouraging, with both types of analysis achieving over 90% accuracy, performance dramatically reduced to 25% when adding a known author text block. This calls into question whether the more accurate results were a function of the software

being biased toward selecting the unknown author. Reiterating the need to explore the various features and analyses available in JGAAP, when such adjustments were made, the software was able to perform at 87.5% in classifying a known text block as "known" when mixed with "other" test blocks.

JStylo performed the best among all three software packages across the range of testing. The user interface was straightforward and the adding of texts was easy to conduct. Just like JGAAP, JStylo has numerous options from which to choose. Although the program performed well, further exploration into the capabilities of various combinations of features and analyses will ensure that any improvements in performance can be identified. Both Naïve Bayes and Simple Logistic Regression analyses outperformed other software during primary testing. With an accuracy of 88.9%, Simple Logistic Regression showed the best promise for potential use in cases of suspected contract cheating. Even more evidence of this was 100% accuracy of Simple Logistic Regression when categorizing a known text block as "known" when mixed with "other" test blocks.

Comparing the use of stylometry software to existing contract cheating detection methods, it appears that the former is superior both in terms of accuracy and resources required. Current methods include text-matching detection software and human evaluators. Current text-matching software unvaryingly has been shown to be unable to highlight ghostwritten texts and new services designed to investigate authorship have yet to be tested by a wide range of users and researchers. Evaluator-based detection is extremely time-consuming and assumes the grader is either familiar with the previous work of the student, has received training about attributes of ghostwritten works, or both. Neither of these necessary defenses is likely to exist on a wide scale in contemporary higher education. In particular, as stated by Singh and Remenyi (2016), familiarity with student writing has become more difficult through growing class sizes and different graders being used across assignments. As noted by Lines (2016), even if contract cheating is suspected by evaluators, one cannot be sure that the cause of the suspicion is because a student's work has genuinely improved or they simply are weak academic writers.

The accuracy of human-based detection has been shown to vary significantly. In Dawson and Sutherland-Smith (2018), graders averaged 57.2% accuracy when classifying contract versus legitimate works and 62% of contract texts were correctly identified. Even after receiving training about catching contract cheating in Dawson and Sutherland-Smith's (2019) follow-up study, among contractor written papers, graders only identified 24.6% correctly although overall accuracy reached 85.7%. Not only can JStylo outperform these human-based procedures, this software would also take substantially less time and effort than a series of manual grading exercises. Moreover, both Rogerson (2017) and Dawson and Sutherland-Smith (2019) admitted that a grader's "hunch" would not be sufficient evidence to support a contract cheating allegation; thus, having additional tools, such as JStylo, are critical to appropriate handling of such occurrences.

While research by Neal et al. (2018) and Brocardo et al. (2015) reported high levels of accuracy in authorial identification, around 95%, these studies used much more sophisticated methods of analysis. Further, these researchers had significant knowledge of computer programming and used models that would require users to have the same level of knowledge. This certainly would require more out of assessors than can be reasonably expected in most cases. Thus, relying on COTS software such as JGAAP and JStylo are more practical options (Juola, 2017).

As with all studies, this research was subject to certain assumptions, limitations, and delimitations. The study was limited by potential variations within the documents randomly chosen to include in the study. An assumption was that since the documents came from the same journal, they were written in the same topic area, had similar tone, and were of comparable sentiment. The findings were also limited due to potential difference in software performance on different types of documents thus, at this point, it cannot be assumed that software accuracy performance will be consistent across types of authors, categories of documents, or genres. The accuracy of stylometry software is potentially limited if an author attempts to obfuscate their work in an attempt to replicate a specific authorial fingerprint, although it seems unlikely that a ghostwriter would have the knowledge, time, or energy to do this for a student. One delimitation was the selection of a specific journal on which to focus this pilot study. Another delimitation was the sample sizes. Although the selected sample sizes followed the guidance of available literature, larger and more diverse samples may have resulted in different findings. Future comprehensive studies are planned to address these weaknesses to provide more robust conclusions.

Although stylometry software may not yet be suitable for a standalone solution to contract cheating, it could, and conceivably should, be part of the evolutionary and systematic approach advocated by Rogerson (2017). Based on the findings of this study as well as those within the literature support, a methodical procedure should be developed for stakeholders to adopt in efforts to curb contract cheating. First, identification of caution "flags" is essential. These include a noticeable change in student writing, oddities in language (e.g., British versus American English), unusually low text similarity values, not answering the question, inappropriate references, misrepresented references, and past academic integrity violations by the student. One or more of these flags may warrant further investigation via one or more COTS stylometry software packages. This would also build a better case for stakeholders prior to confronting a student. Lastly, the knowledge by students that there are computer-based contract cheating detection methods in use may act as a deterrent.

Conclusion

The purpose of this pilot study was to determine if stylometry software could be a potential solution to the growing problem of online contract cheating. Based on the findings of this study, it is apparent that such software is capable of accurately detecting anomalies in authorship among documents. Just as text-matching detection software is not infallible and requires some level of interpretation, stylometry analysis cannot be expected to provide 100% accuracy and may require further appraisal by graders. Yet stylometry appears to provide the most auspicious solution to the quandary of ghostwritten assignments available. Further, the ease of use by computer novices as well as the lack of cost indicates that COTS stylometry software can provide a practical and low-cost answer to the question of how to detect and deter contract cheating. Moreover, upon familiarizing oneself with these software packages, testing could be completed within minutes in lieu of more tedious sleuthing through student papers.

What is clear is that stakeholders are losing the contract cheating battle and can no longer stand idle in the face of the issue. With further research into the best practices in terms of features, processing, and analytics called upon by the software can provide even more accuracy. As part of a systematic approach to contract cheating, stylometry software gives stakeholders a semblance of hope in dealing with this mounting threat to academic integrity. Just as importantly to catching

contract cheaters, stylometry provides a means of avoiding accusing students who submit legitimate work of something they have not done. In sum, stylometry software is an obvious and competent means of addressing contract cheating.

Recommendations for Future Research

Based on the finding of this study coupled with exigent literature, the following recommendations are made for future research. Researchers are encouraged to conduct a larger study utilizing JGAAP and JStylo to identify the best practices of software settings to improve accuracy and reliability. It is also recommended that a study takes place assessing stylometry software using actual student work versus contracted documents that address the same question or task. Lastly, studies on new contract cheating services, such as Turnitin's Authorship Investigate, should be undertaken to assess their capabilities.

References

Anekwe, T. D. (2010). Profits and plagiarism: the case of medical ghostwriting. *Bioethics*, *24*(6), 267–272. doi: 10.1111/j.1467-8519.2008.00705.x

Athira, U., & Thampi, S. M. (2018). An author-specific-model-based authorship analysis using psycholinguistic aspects and style word patterns. *Journal of Intelligent & Fuzzy Systems*, *34*(3), 1453–1466. doi: 10.3233/JIFS-169441

Bailey, J. (2019, May 14). *PayPal's essay mill ban has had no impact*. Plagiarism Today. https://www.plagiarismtoday.com/2019/05/14/paypals-essay-mill-ban-has-had-no-impact/

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., ... & van Haeringen, K. (2018). Contract cheating: A survey of Australian university students. *Studies in Higher Education*, 1–20. doi: 10.1080/03075079.2018.1462788

Brocardo, M. L., Traore, I., & Woungang, I. (2015). Authorship verification of e-mail and tweet messages applied for continuous authentication. *Journal of Computer and System Sciences*, *81*(8), 1429–1440. doi: 10.1016/j.jcss.2014.12.019

Clare, J., Walker, S., & Hobson, J. (2017). Can we detect contract cheating using existing assessment data? Applying crime prevention theory to an academic integrity issue. *International Journal for Educational Integrity*, *13*(1), 4. doi: 10.1007/s40979-017-0015-4

Coughlan, S. (2019, April 3). *PayPal to reject essay-writing firms*. BBC. https://www.bbc.com/news/education-47800531

Dawes, B., Merivale, H., & Millican, J. (2003). *Signature Stylometry System 1.0* [Computer software]. PhiloComp. http://www.philocomp.net/humanities/signature.htm

Dawson, P., & Sutherland-Smith, W. (2018). Can markers detect contract cheating? Results from a pilot study. *Assessment & Evaluation in Higher Education*, *43*(2), 286–293. doi:10.1080/02602938.2017.1336746

Dawson, P., & Sutherland-Smith, W. (2019). Can training improve marker accuracy at detecting contract cheating? A multi-disciplinary pre-post study. *Assessment & Evaluation in Higher Education*, *44*(5), 715–725. doi: 10.1080/02602938.2018.1531109

EVL Labs. (2018). *Java Graphical Authorship Attribution Program* [Computer software]. https://evllabs.com/?page_id=42

Farhi, P. (2014). Who wrote that political memoir? No, who actually wrote it? *The Washington Post.* https://www.washingtonpost.com/lifestyle/style/who-wrote-that-political-memoir-no-who-actually-wrote-it/2014/06/09/

García, A. M., & Martín, J. C. (2012). Testing Delta on the "Disputed Federalist Papers." *International Journal of English Studies*, *12*(2), 133–150. doi:10.6018/ijes/2012/2/161791

Harper, R., Bretag, T., Ellis, C., Newton, P., Rozenberg, P., Saddiqui, S., & van Haeringen, K. (2018). Contract cheating: A survey of Australian university staff. *Studies in Higher Education*, *44*(11), 1–17. doi: 10.1080/03075079.2018.1462789

Juola, P. (2017). Detecting contract cheating via stylometry methods. *Plagiarism Across Europe and Beyond 2017*, 187–198.

Juola, P., Sofko, J., & Brennan, P. (2006). A prototype for authorship attribution studies. *Literary and Linguistic Computing*, *21*(2), 169–178.

Lancaster, T. (2019). Profiling the international academic ghost writers who are providing low cost essays and assignments for the contract cheating industry. *Journal of Information, Communication and Ethics in Society*, *17*(1), 72–86. doi: 10.1108/JICES-04-2018-0040

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

Ma, H. J., Wan, G., & Lu, E. Y. (2008). Digital cheating and plagiarism in schools. *Theory Into Practice*, *47*(3), 197–203. doi: 10.1080/00405840802153809

Medway, D., Roper, S., & Gillooly, L. (2018). Contract cheating in UK higher education: A covert investigation of essay mills. *British Educational Research Journal*, *44*(3), 393–418. doi:10.1002/berj.3335

Millican, P. (2003). The Signature Stylometry System: A user-friendly system for textual analysis. *The Leeds Electronic Text Centre.* http://www.philocomp.net/humanities/signature.htm

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem: A comparative study of discrimination methods applied to the authorship of the disputed Federalist Papers. *Journal of the American Statistical Association*, *58*(302), 275–309.

National Center for Education Statistics. (2019). *Fast facts*. IES. https://nces.ed.gov/fastfacts/display.asp?id=372

Neal, T., Sundararajan, K., Fatima, A., Yan, Y., Xiang, Y., & Woodard, D. (2018). Surveying stylometry techniques and applications. *ACM Computing Surveys (CSUR)*, *50*(6), 86. doi:10.1145/3132039

Newton, P. (2018). How common is commercial contract cheating in higher education? *Frontiers in Education, 3*, 1–18. doi: 10.3389/fedue.2018.00067

Nieto, V. G., Sierra, C. V., Juan, M. P., Barco, P. M., & Cueto, A. S. (2008). Exploring state-of-the-art software for forensic authorship identification. *International Journal of English Studies*, *8*(1), 1–28.

Perkins, M., Gezgin, U. B., & Roe, J. (2018). Understanding the relationship between language ability and plagiarism in non-native English speaking business students. *Journal of Academic Ethics*, *16*(4), 317–328.

Posadas-Durán, J. P., Gómez-Adorno, H., Sidorov, G., Batyrshin, I., Pinto, D., & Chanona Hernández, L. (2017). Application of the distributed document representation in the authorship attribution task for small corpora. *Soft Computing*, *21*(3), 627–639. doi:10.1007/s00500-016-2446-x

Prasad, S. N., Narsimha, V. B., Reddy, P. V., & Babu, A. V. (2015). Influence of lexical, syntactic and structural features and their combination on authorship attribution for Telugu text. *Procedia Computer Science*, *48*, 58–64. doi: 10.1016/j.procs.2015.04.110

Puig, X., Font, M., & Ginebra, J. (2016). A unified approach to authorship attribution and verification. *The American Statistician*, *70*(3), 232–242. doi:10.1080/00031305.2016.1148630

Ragel, R., Herath, P., & Senanayake, U. (2013, December). Authorship detection of SMS messages using unigrams. In *2013 IEEE 8th International Conference on Industrial and Information Systems* (pp. 387–392). IEEE.

Rigby, D., Burton, M., Balcombe, K., Bateman, I., & Mulatu, A. (2015). Contract cheating & the market in essays. *Journal of Economic Behavior & Organization*, *111*, 23–37. doi:10.1016/j.jebo.2014.12.019

Rogerson, A. (2014). Detecting the work of essay mills and file swapping sites: Some clues they leave behind. In *6th International Integrity & Plagiarism Conference: Conference Proceedings* (pp. 1–9). Plagiarismadvice.org.

Rogerson, A. M. (2017). Detecting contract cheating in essay and report submissions: Process, patterns, clues and conversations. *International Journal for Educational Integrity*, *13*(1), 10. doi: 10.1007/s40979-017-0021-6

Salkind, N. J. (2012). *Exploring research* (8th ed.). Pearson.

Sarwar, R., Li, Q., Rakthanmanon, T., & Nutanong, S. (2018). A scalable framework for cross lingual authorship identification. *Information Sciences*, *465*, 323–339. doi:10.1016/j.ins.2018.07.009

Singh, S., & Remenyi, D. (2016). Plagiarism and ghostwriting: The rise in academic misconduct. *South African Journal of Science*, *112*(5–6), 1–7. doi:10.17159/sajs.2016/20150300

Singleton, P., & Ricksen, M. (2019, April). *Your guide to recognizing various forms of plagiarism and gaining insights from technology*. University of Calgary, Canada. https://prism.ucalgary.ca/handle/1880/110294

Slade, C., Rowland, S., & McGrath, D. (2019). Talking about contract cheating: facilitating a forum for collaborative development of assessment practices to combat student dishonesty. *International Journal for Academic Development*, *24*(1), 21–34.

Smith, T. (2019, April 10). *Buying college essays is now easier than ever. But buyer beware*. NPR. https://www.npr.org/2019/04/10/710985799/contract-cheating-colleges-crack-down-on-ghostwritten-essays

Stolerman, A., & Dutko, T. (2013). *JStylo Authorship Attribution Framework v1.2* [Computer software]. Drexel University. https://psal.cs.drexel.edu/index.php/JStylo-Anonymouth

Taylor, S. M. (2014). Term papers for hire: How to deter academic dishonesty. *The Education Digest*, *80*(2), 52.

Turnitin. (2019). *Authorship Investigate*. https://www.turnitin.com/products/authorship-investigate

Top 20 Writing Services. (2019). *The 20 best dissertation writing services in 2019*. https://top20writingservices.com/best-dissertation-writing-services/

Wallace, M. J., & Newton, P. M. (2014). Turnaround time and market capacity in contract cheating. *Educational Studies*, *40*(2), 233–236. doi:10.1080/03055698.2014.889597

White, J. L. (2016). Shadow Scholars and the Rise of the Dissertation Service Industry: Can We Maintain Academic Integrity? *Journal of Research Practice*, *12*(1), 1–9.

Appendix A: Figures

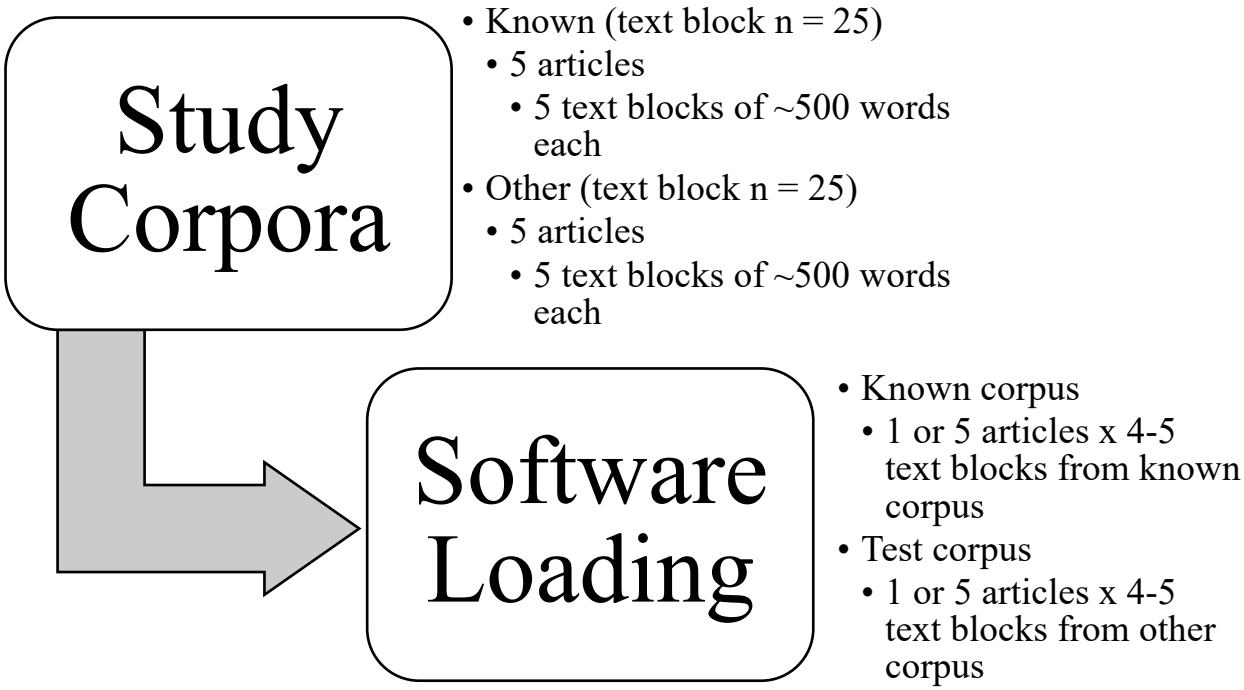

Figure 1. Document selection process for SSS.

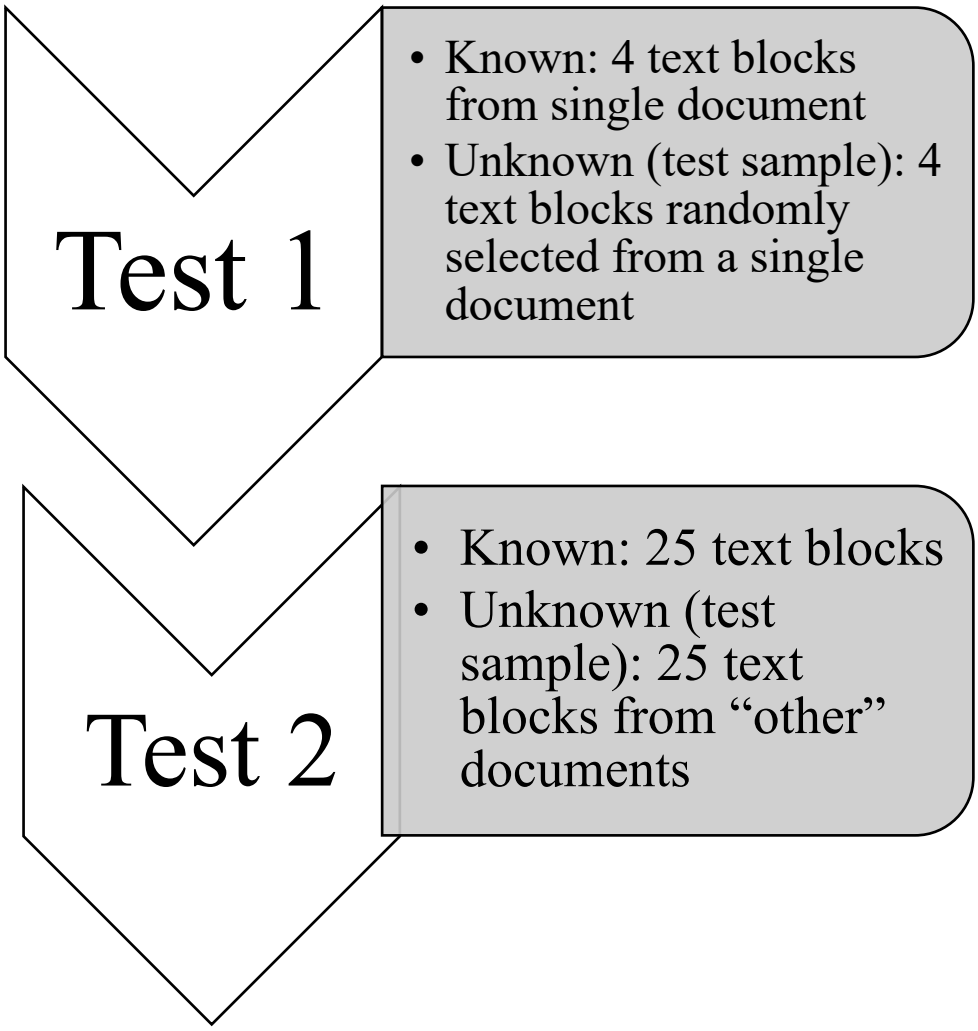

Figure 2. Testing procedures for SSS.

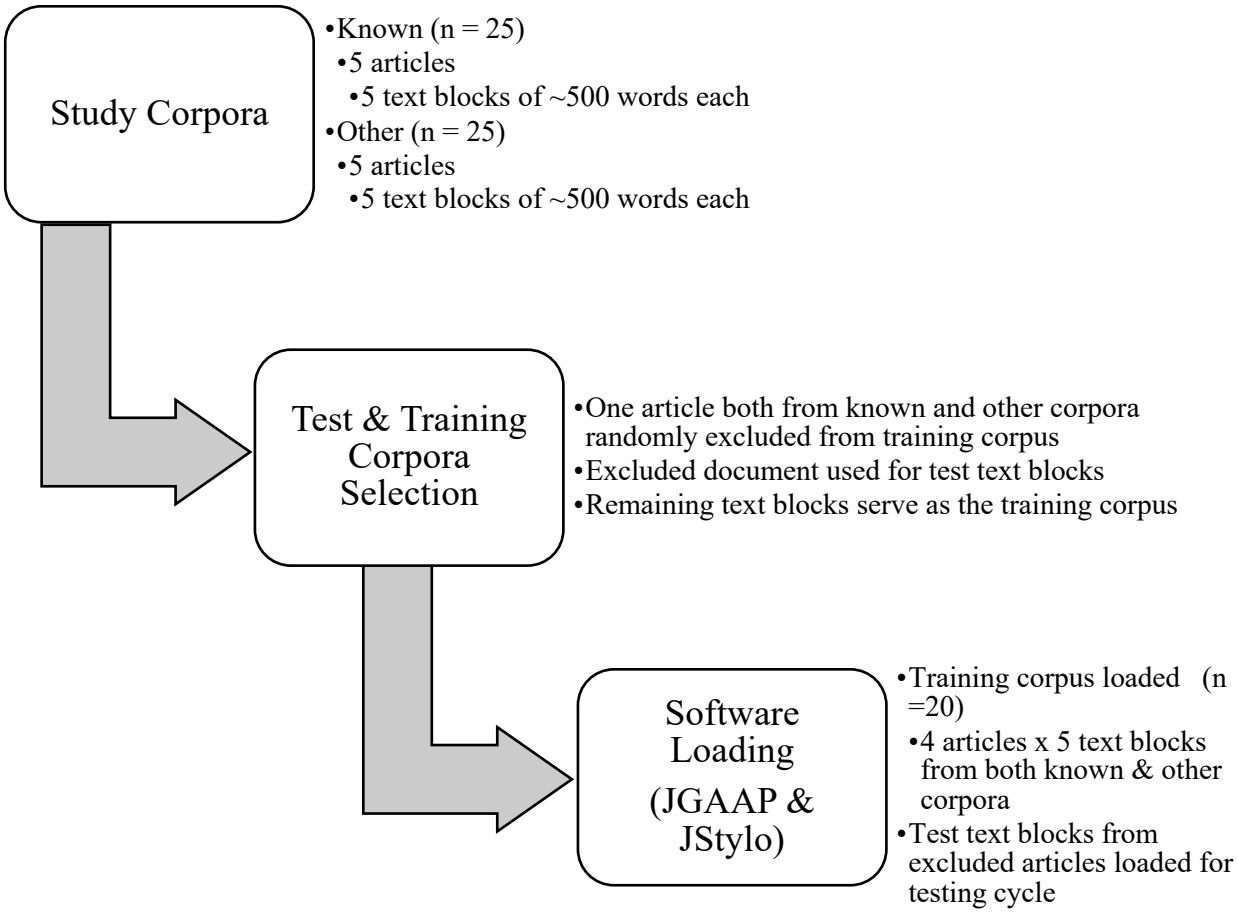

Figure 3. Document selection process for JGAAP and JStylo.

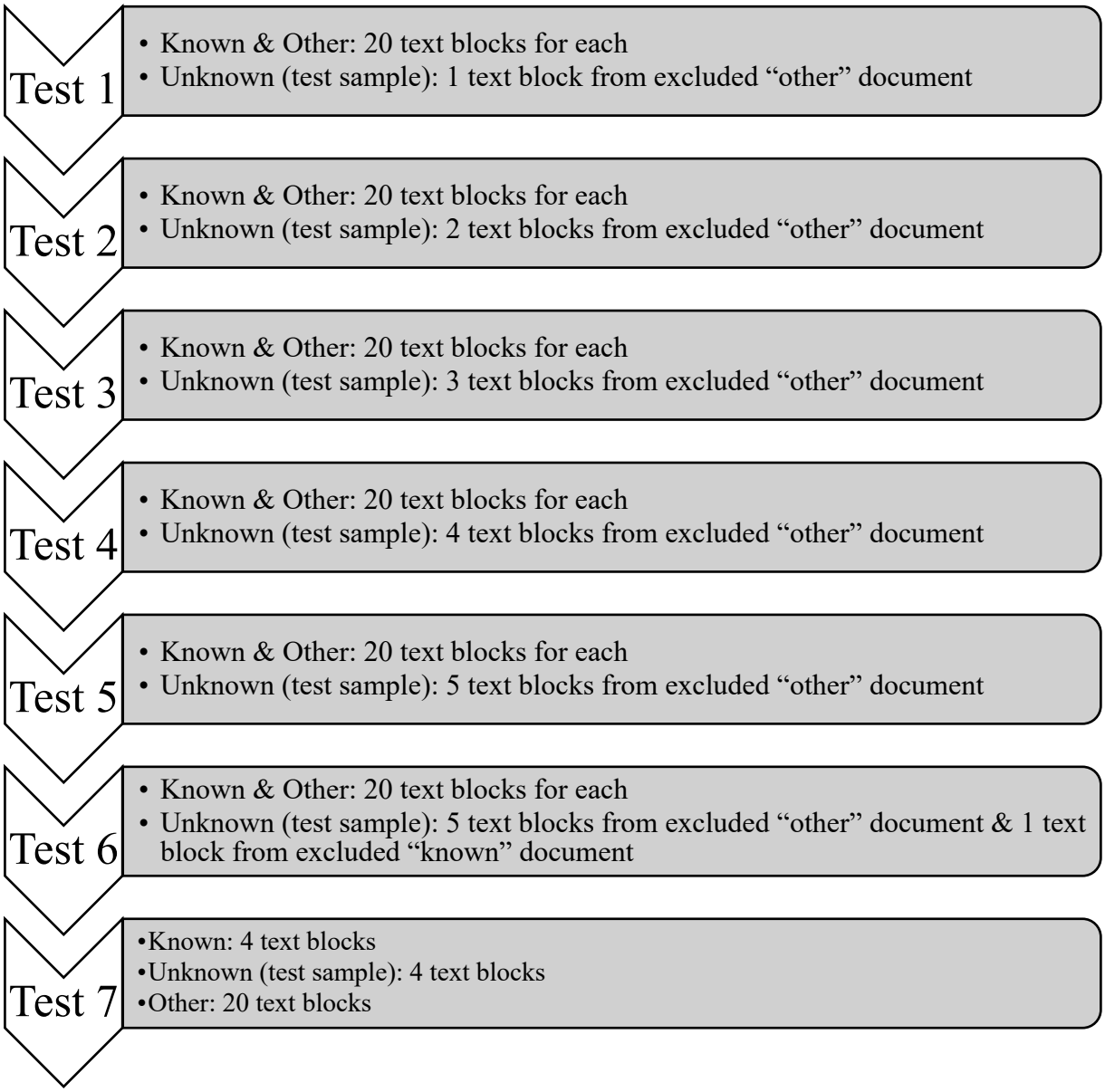

Figure 4. Testing procedures for JGAAP and JStylo.