Inquiry: The Journal of the Virginia Community Colleges

Volume 19
Number 1 *2014*

Article 6

2014

A Comparative Analysis of SafeAssign and Turnitin

Jared Hunt MA

Patrick Tompkins PhD

A Comparative Analysis of SafeAssign and Turnitin

JARED HUNT, MA

PATRICK TOMPKINS, PHD

With the proliferation of the internet and online sources, student plagiarism—either purposeful or accidental—is a prevalent issue at colleges in the 21st century. In the ideal classroom, plagiarism would be a non-issue, and instructors and students could dedicate both their time and energy wholly to the exercises of teaching and learning. This ideal, however, is far from reality, making plagiarism detection software an integral part of an instructor's available pedagogical tools. Unfortunately, both SafeAssign and Turnitin, the two plagiarism detection programs analyzed in this report, are imperfect tools. As Gillis, Lang, Norris, and Palmer (2009) have noted, "Despite the verbiage on the applications' Web sites to the contrary, nothing about the interfaces suggests an emphasis on teaching or learning about proper citation methods" (p. 60).

The results of this study should help educators determine if either program is superior to the other and which program best justifies the allocation of technology resources for plagiarism prevention.

Despite this rather pessimistic view, both SafeAssign, which comes integrated with Blackboard course management system in the Virginia Community College System (VCCS), and Turnitin, the most popular plagiarism detection program (Stapleton, 2012), may be viable options for detecting student plagiarism. As instructors in Virginia's Community Colleges, we have found that some faculty prefer Turnitin to SafeAssign even though SafeAssign is seamlessly integrated into Blackboard and is therefore available to VCCS faculty at no additional cost. These faculty members believe that Turnitin detects more plagiarism and is more intuitively designed and easier to navigate. This impression, however, appears to be conjectural, anecdotal, and/or subjective. In reality, very few comparative studies are available to support the general impression of either program's superiority. The most recent publicly-available studies—the Gillis, et al. (2009) examination of 400 first-year student texts from Texas Tech University and Wise, Kim, Stierwalt, and Hu's (2009) Florida State University study of sets of original and fabricated texts—both date from 2009. Since then, both SafeAssign and Turnitin have upgraded their software and databases.

To evaluate SafeAssign and Turnitin further, we recently processed a sample group of student texts from multiple disciplines and classes at one of Virginia's

Community Colleges through both programs. The purpose of the study was to assess which program (1) detected more instances of plagiarism, (2) mistakenly identified more false positives, and (3) was more intuitive and/or easier to use. The results of this study should help educators determine if either program is superior to the other and which program best justifies the allocation of technology resources for plagiarism prevention.

LITERATURE REVIEW

Both professional and popular literature suggests that we are experiencing an unparalleled era of plagiarism, abetted by technology resources that provide access to billions of source materials to anyone with internet access (Evans, 2006; Hansen, 2003). Concomitantly, the availability of anytime, anywhere sources may contribute to relaxed attitudes or naiveté about the fair use of others' ideas and texts (Jiang & McKauge, 2013). On the other hand, the proliferation of technology as well as its ease of use and increasing sophistication has also provided instructors with greater opportunities to prevent and detect plagiarism (Purdy, 2005; Wise, Kim, Stierwalt, & Hu, 2009). In addition to Google, the most ubiquitous enabler of both plagiarism violation and detection (Maurer, Kappe, & Zaka, 2006; Purdy, 2005), a number of software programs have been developed to assist instructors' efforts to ensure academic integrity, including both SafeAssign and Turnitin. However, educators have raised concern about both the effectiveness and appropriateness of these two programs (Gillis, Lang, Norris, & Palmer, 2009; Howard, 2003). Some evidence suggests that the programs return a disproportionate number of false positive hits; that is, passages are flagged as potentially plagiarized content, but they are simply phrases commonly used in English writing or the results of mechanical or stylistic errors with paraphrasing and quoting (Stapleton, 2012).

Scant research has been published on the comparative effectiveness of different software programs to accurately detect plagiarism in student writing. In their study at Florida State University, Wise, et al. (2009) reported that the two programs were generally equally effective and recommended them as acceptable resources for instructors, although SafeAssign performed better when detecting potential plagiarism with paraphrased materials. Gillis, et al. (2009) found that the programs were statistically significantly different in their reports of potentially plagiarized passages, but the authors did not further evaluate which program was more accurate. Both teams of researchers reported that while Turnitin identified more sources of plagiarized material in a student's text, this higher number was the result, primarily, of multiple hits for a potentially plagiarized passage rather than from multiple instances of plagiarism.

ACCESSING AND USING THE PROGRAMS

The VCCS has purchased a system-wide license for the Blackboard course management system, the cost of which is shared among Virginia's 23 Community Colleges. SafeAssign is seamlessly integrated into Blackboard (it can be accessed directly through the course menu), and it is available at no additional cost to colleges, instructors, or students. On the other hand, if instructors wish to use Turnitin, their college must purchase a separate license. Although costs for specific

licenses vary among institutions, at one VCCS college with an annual full-time equivalent student (FTES) count of over 6,000, the current annual price for Turnitin is roughly $22,000.

Both programs work by comparing a submitted student text to a series of continuously updated databases which draw on texts from publically accessible sites on the internet, academic journals, and all papers previously submitted to either program.

Instructors can submit texts to SafeAssign through a "DirectSubmit" tab; they can submit texts to Turnitin through a "Submit paper" tab. Although these features are very similar, Turnitin requires the creation of "class" and "assignment" folders because it is not integrated with Blackboard. Instructors can also direct students to submit their own texts to SafeAssign or Turnitin. Again, these features are also very similar. However, Turnitin requires more steps by both instructors and students because it is not an embedded Blackboard feature. After the texts have been submitted, both programs create a report that identifies passages that may have been drawn from other sources, whether legitimately or illegitimately, intentionally or unintentionally. Both programs report plagiarism as a percentage of the total text submitted, and highlight text for which a match has been found. The determination of plagiarism is left to the instructor who must review the passages flagged in the student's text and compare that flagged content against the source material procured by the programs. Turnitin has a slight advantage in this area, in that it has two options SafeAssign does not possess. Instructors who use Turnitin can choose to "Exclude Quotes" and/or "Exclude Bibliography." These options are imperfect in Turnitin (correctly cited quotations and bibliographic information are often still flagged as matching other source material and therefore potentially plagiarized), but they are nonexistent in SafeAssign; these options lower the number of false positives reported in the former program.

RESEARCH QUESTIONS

This study sought to answer the general question of whether the SafeAssign or the Turnitin plagiarism prevention program was generally superior. The specific research questions follow:

1) Which plagiarism detection software (SafeAssign or Turnitin) is more effective at detecting plagiarism in student texts?
    a) Does SafeAssign or Turnitin detect more instances of actual plagiarism?
    b) Does SafeAssign or Turnitin report more false positives (instances of flagged plagiarism that are not actually plagiarized)?
2) Which plagiarism detection software (SafeAssign or Turnitin) is easier for faculty to use?
    a) Which detection software has a more intuitive design?
    b) Which detection software requires fewer steps on the part of both the instructor and the students while in use?

Parameters

In order to address the main research question about which program was more effective at detecting plagiarism, it was first necessary to set the parameters for plagiarism, which also entailed defining what was not plagiarism. Text was considered plagiarized if it was found to match text from either program's bank of databases and if that match was determined not to be a false positive.

False positives, or matches between the student text and the database text that most likely did not represent instances of intentional plagiarism, were determined on a case-by-case basis. Four categories of false positives were identified by the researchers, and if the highlighted text was determined to belong to one of these four categories, it was noted as a false positive. The false positive types were the following:

- Quotations and Paraphrases: if language in quotation marks was flagged by SafeAssign or Turnitin as potentially plagiarized, it was recorded as a false positive. If text was improperly paraphrased (due to similar vocabulary and sentence structure) but there was in-text documentation, it was recorded as a false positive based on the assumption that, if the student indicated his or her source with in-text documentation, the occurrence was more likely to be a case of incorrect paraphrasing rather than purposeful plagiarism.
- Bibliography/Works Cited: if language in a bibliography or a works cited page was flagged by SafeAssign or Turnitin as potentially plagiarized, it was recorded as a false positive.
- Fractured Language: if in a sentence only some of the smaller phrases and words were flagged by SafeAssign or Turnitin as potentially plagiarized, it was counted as a false positive. For example, in the sentence "We like to think that only bad people do bad things, and that good people never do bad things," the disjointed phrases "that only bad people do bad things," "good people," and "do bad things" were flagged by Turnitin as potentially plagiarized, while the connecting text was not.
- Common Language: if language that is commonly used by many sources—such as clichés, a book title, a famous athlete's name, or a common phrase—was flagged by SafeAssign or Turnitin as potentially plagiarized, it was counted as a false positive. Like fractured language, these matches tended to show up in sentences in which only one or two percent of the text was flagged by SafeAssign or Turnitin. A typical example of common language is the sentence "I do not think that I could get a loan at this rate," which was flagged by SafeAssign because it found "I do not think that I could ever get a girlfriend, anyway" (63% matching) from the website Answers.yahoo.com.

For each passage in a student's text that was found to match text existing in the programs' databases, the flagged phrases, sentences, or paragraphs, as highlighted by the program, were evaluated using these four categories to determine if the matching text was or was not a case of actual plagiarism.

Methodology

In order to compare the programs, a total of 293 texts were collected from courses at a Virginia community college or were created by the lead researcher as controls. Students whose work was used in the study signed informed consent forms. Student identifying information was removed from each text prior to submission for analysis within the plagiarism detection programs, and each text was assigned a unique identifier that could not be traced back to the student but that allowed for tracking within the study. Each text was run through both SafeAssign and Turnitin.

Both programs have an option to exclude specific sources. In SafeAssign this option was implemented for any source which referred to "Another Student's Paper" or the "VCCS." In Turnitin this option was implemented for any source which referred to the college from which the papers were collected. This step ensured that no text would test positive for potentially plagiarized passages because it had previously been submitted by a professor during the normal course of the semester or because it had been submitted earlier during the course of this research.

For each text, three data points were recorded for both SafeAssign and Turnitin: (1) the percentage of the text where matches were found by the program, (2) of that percentage, the amount that was determined by the lead researcher to be false positives, and (3) a classification by the lead researcher of each false positive into one of the four false positive categories described in the parameters section above. Additionally, the texts were run through Turnitin a second time using the option to "Exclude Quotes" and "Exclude Bibliography" (as previously stated, only Turnitin had this option), and the same three data points noted above were recorded.

Population/Sample

Texts were solicited from students enrolled in first-year courses at one of Virginia's Community Colleges during Fall 2012 and Spring 2013 semesters. A total of 293 texts were collected for the study (Table 1).

Of these, 194 texts were collected from English composition, literature, and religion courses, while a total of 54 were collected from psychology and mathematics courses. In addition, 33 autobiographies—assignments which, due to the nature of the writing, were unlikely to be accidentally or purposefully plagiarized—were collected from an English composition class during Fall 2012 in order to act as non-plagiarized controls. Finally, the lead researcher assembled the entirety of 12 texts from various "free essay" websites to serve as plagiarized controls. Nine texts, some of which were duplicate submissions and some of which were texts that yielded error messages when submitted to Safe Assign or Turnitin, were eliminated from the sample, resulting in 284 valid texts for analysis.

Table 1: Sources of Texts for the Study

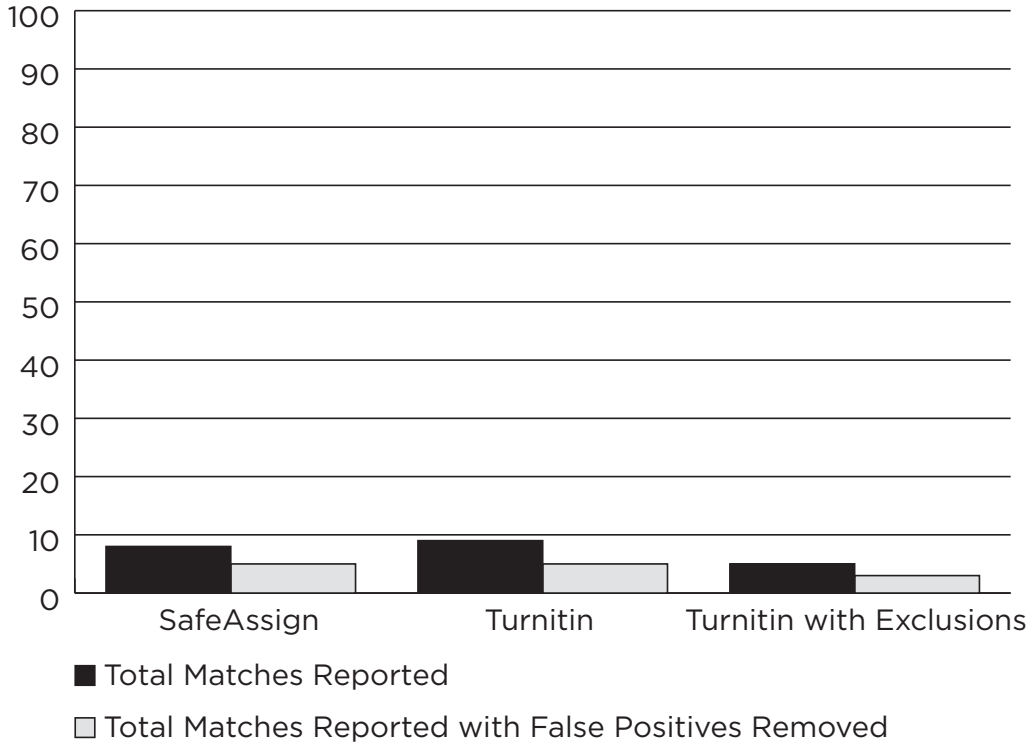
English / religion courses	Psychology / math courses	Autobiographies	Plagiarized controls
194	54	33	12

Data Collection, Analysis, and Findings

The 284 valid texts were analyzed to test the research questions. Each text was submitted to the SafeAssign and Turnitin programs and analyzed for content that was found to match content existing in the programs' databases. The unfiltered number of "hits," indicated as a percentage of the total text, was recorded for each text. After false positives were eliminated using the procedures described above, the two numbers were contrasted to determine the percentage of text that was found to match text existing in the programs' databases but not fitting the criteria determining a false positive. An analysis of the data revealed the following average (mean) statistics which are also visually represented in Figure 1:

- SafeAssign: 6.95% of all text was found to match text existing in the program's databases, which included an average of 4.2% of the text that was found to be false positives. Thus, on average, SafeAssign identified 2.75% of student text as instances of actual plagiarism.
- Turnitin: 7.64% of all text was found to match text existing in the program's databases, which included an average of 4.26% of the text that was found to be false positives. Thus, on average, Turnitin identified 3.38% of student text as instances of actual plagiarism.
- Turnitin using the "Excludes Quotes" and "Exclude Bibliography" option: 4.51% of all text was found to match text existing in the program's databases, which included an average of 1.33% of the text that was found to be false positives. Thus, on average, Turnitin using the option to exclude quotes and bibliography identified 3.18% of student text as instances of actual plagiarism.

Figure 1. SafeAssign and Turnitin Plagiarism Matching Reports for 284 Texts as Percent of the Total Text



Descriptive statistics indicated that the sample was significantly leptokurtic and skewed right, indicating non-normal distribution, due to the preponderance of texts in which no instances of plagiarism were identified, both with and without screening out false positives and even when the autobiographies, which were assumed a priori to not contain plagiarized passages, were not included in the data set. The application of transformations to the data, such as square root and log transformations, were unsuccessful in inducing normality.

For a study of this kind—in which the predictor variable (SafeAssign versus Turnitin) is categorical, the outcome variable (percentage of plagiarized passages) is continuous, and the same subjects (student texts) are used in each condition—the dependent samples T-Test is the appropriate statistical methodology (Field, 2009). However, because the non-normally distributed data violated the assumptions of the T-test, the Wilcoxon signed rank test is appropriate and was therefore used in this study (Field, 2009).

For the 284 valid texts, there was no statistically significant difference in plagiarism detected by Turnitin ($Mdn = 0.00$) compared to SafeAssign ($Mdn = 1.00$), $Z = -.251$, $p > .05$. For the 284 valid texts, there was no statistically significant difference in false positives detected by Turnitin ($Mdn = 0.00$) compared to SafeAssign ($Mdn = 0.00$), $Z = -.854$, $p > .05$.

After removing false positives, the lead researcher determined that only 31 of the 284 texts, including the 12 plagiarized controls, contained actual instances of plagiarism. When the SafeAssign and Turnitin reports were statistically analyzed for these 31 texts, plagiarism detected was statistically significantly higher for Turnitin ($Mdn = 37.00$) compared to SafeAssign ($Mdn = 34.00$), $Z = -2.058$, $p < ,05$, but the effect size was small ($r = -.122$).

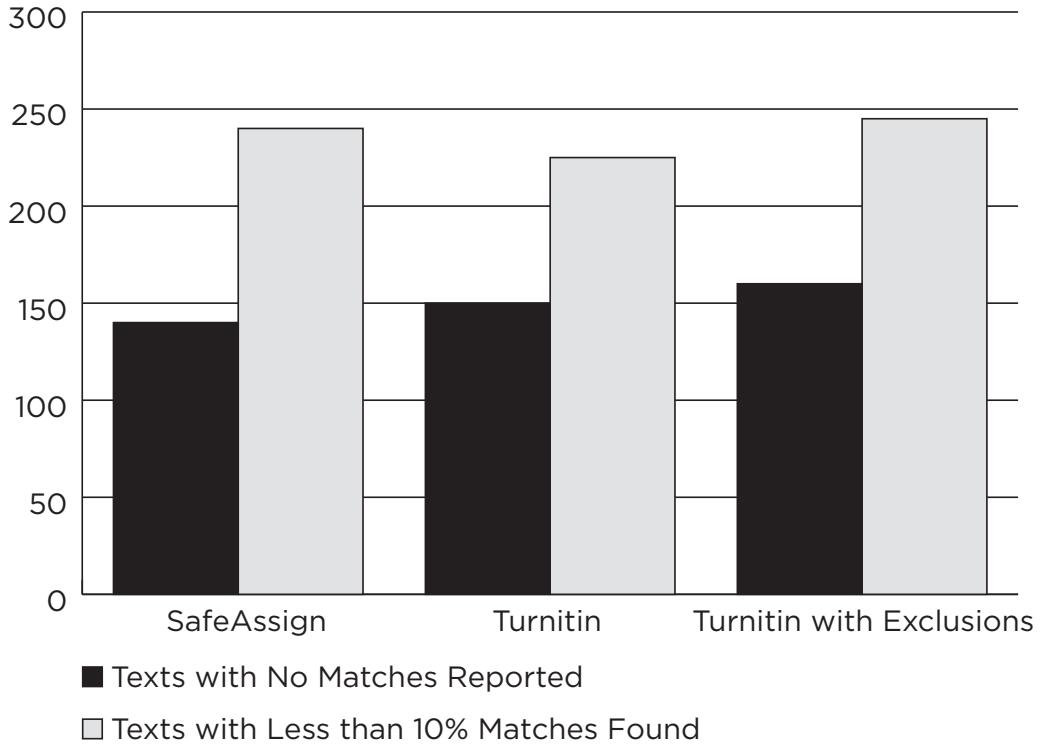
In sum, little meaningful difference was apparent in the ability and accuracy of SafeAssign and Turnitin to identify plagiarism: both programs grossly over-reported false positives and often failed to identify actual, blatant cases of plagiarism.

Other revealing numbers can be gathered from the data. For example, the programs performed similarly when reporting on texts in which no or few instances were reported of passages matching text found in the programs' databases (Figure 2):

- SafeAssign reported no matches in 129 texts (44% of the total) and matching passages of 10% or less in 235 texts (80.2% of the total).
- Turnitin reported no matches in 148 texts (50.5% of the total) and matching passages of 10% or less in 223 texts (76.1% of the total).
- Turnitin with "Exclude Quotes" and "Exclude Bibliography" marked reported no matches in 168 texts (57.3% of the total) and matching passages of 10% or less in 261 texts (89.1% of the total).

Of importance here is that in the vast majority of student texts gathered for this research, the occurrence of plagiarism appeared to be exceedingly small.

Figure 2. Distribution of 284 texts where SafeAssign and Turnitin Reported No Matching Text and Less Than 10% Matching Text.



Finally, both programs performed similarly, and did a remarkably poor job, in detecting blatant cases of plagiarism in the 12 plagiarized controls which were created by using "free essay" internet sites (Table 2). Out of the 12 plagiarized control texts, in more than half of the sample (N = 7) SafeAssign found matches for less than 50% of the text, and in one third of the sample (N = 4) SafeAssign found matches for less than 30% of the text. For Turnitin, in half of the sample (N = 6) matches were found for less than 50% of the text, and in one quarter of the sample (N = 3) matches were found for less than 30% of the text.

Table 2: Percentage of Passages in 12 Plagiarized Control Texts Where SafeAssign and Turnitin Reported Matches with Text in their Databases

Plagiarized Control	SafeAssign	Turnitin
Text 1	100%	100%
Text 2	100%	100%
Text 3	100%	100%
Text 4	100%	99%
Text 5	100%	99%
Text 6	44%	98%
Text 7	40%	49%
Text 8	29%	37%
Text 9	21%	34%
Text 10	11%	25%
Text 11	10%	21%
Text 12	8%	4%

Discussion

This study indicated no meaningful difference between SafeAssign and Turnitin in either the effectiveness of matching plagiarized student texts to texts from the programs' databases or in avoiding false positives while doing so. Furthermore, as noted earlier, both programs did a remarkably poor job in detecting the text that was deliberately plagiarized from online sources by the lead researcher as a control.

Also of note is the fact that the number of student texts in which no matches were detected (44%-57.3% of the sample) or in which less than 10% of matching passages were detected (76.1%-89.1% of the sample) was much higher than expected. Although this result might contradict the conventional wisdom that plagiarism is running rampant throughout academia, this information does accord with findings by Gillis, et al. (2009) that the incidence of plagiarism in college composition courses was low, as well as findings by Stapleton (2012) that most incidents of plagiarism in a graduate-level English as a foreign language (EFL) course were unintentional.

Implications for Practice and Further Research

The results above suggest a number of implications for practice: (1) The importance of detecting plagiarism on most assignments—or of paying for plagiarism detection software to use on most assignments—may not be as critical as anecdotal evidence sometimes suggests. A better allocation of resources (both in terms of faculty time and institutional budgets) may be in developing teaching and learning programs to help students learn paraphrasing and citation skills. (2) The research suggests that the additional purchase of Turnitin is not a productive use of institutional dollars. Because there is no statistically significant difference in the ability of SafeAssign and Turnitin to report plagiarized text or to avoid false positives, the provision of SafeAssign through Blackboard provides instructors with a sufficient plagiarism detection program. (3) If an institution does decide to retain Turnitin software, Turnitin should be integrated as a building block in Blackboard to make it easier for faculty and students to use. (4) Faculty may want to exercise caution in having students submit their own work to SafeAssign and Turnitin because the danger of false positives may engender panic and paranoia; the programs' tendency to flag a disproportionately high percentage of passages in student texts is a matter of concern. Gillis, et al. (2009) note that if students see their work flagged for potential plagiarism without analyzing the results in detail to throw out false positives, they may become discouraged or "may shift from writing to an appropriate human audience to 'writing to the software'" (p. 60).

Further research with more texts—particularly more texts with actual plagiarism—is also recommended. First, because the data set was relatively small and did not include very many texts with significant instances of plagiarism, studies using larger samples of student texts would contribute to the assessment of the effectiveness of the two programs. Furthermore, more study is needed of the comparative effectiveness of the programs for different disciplines. For example, Kaner and Fiedler (2008) found that Turnitin and MyDropBox were unable to detect instances in which students blatantly plagiarized from material published in journals and databases important in the field of software engineering; therefore, it would be useful to include texts from a broader array of traditional liberal arts, STEM-H (science, technology, engineering, math and health professions), and professional and technical disciplines, such as business and information technology courses.

REFERENCES

Evans, R. (2006). Evaluating an electronic plagiarism detection service. *Active Learning In Higher Education,* 7(1), 87-99. doi:10.1177/1469787406061150

Gillis, K., Lang, S., Norris, M., & Palmer, L. (2009). Electronic plagiarism checkers: Barriers to developing an academic voice. *WAC Journal,* 20, 51-62.

Hansen, B. (2003). Combating plagiarism. *CQ Researcher,* 13(32), 775-782.

Howard, R. (2003). Should educators use commercial services to combat plagiarism?. *CQ Researcher,* 13(32), 789.

Jiang, H., Emmerton, L., & McKauge, L. (2013). Academic integrity and plagiarism: A review of the influences and risk situations for health students. *Higher Education Research & Development,* 32(3), 369-380. doi:10.1080/07294360.2012.687362

Kaner, C., & Fiedler, R. L. (2008). A cautionary note on checking software engineering papers for plagiarism. IEEE *Transactions on Education,* 51(2), 184-188. doi:10.1109/TE.2007.909351

Maurer, H., Kappe,m F., & Zaka, B. (2006). Plagiarism—a survey. *Journal of Universal Computer Science,* 12(8), 1050-1084.

Stapleton, P. (2012). Gauging the Effectiveness of Anti-Plagiarism Software: An Empirical Study of Second Language Graduate Writers. *Journal of English for Academic Purposes,* 11(2), 125-133.

Wise, J-M., Kim, A., Stierwalt, K., & Hu, H. (2009). *SafeAssign transition report.* Retrieved from http://ctl.fsu.edu/explore/bestpractices/docs/safeassigntransitionreportfinal.pdf

ABOUT THE AUTHORS

Jared Hunt, M.A., is Adjunct English Instructor at John Tyler Community College in Midlothian, Virginia

Patrick Tompkins, Ph.D., is Dean of Communications, Humanities, and Social Sciences at Thomas Nelson Community College in Hampton, Virginia.