

Developing online materials to teach how to avoid plagiarism, and a database of student essays to check for plagiarism

Joseph V. Dias, English Department, Aoyama Gakuin University

1. Introduction

Thanks to a generous grant from the Information Media Center of Aoyama Gakuin University, the development of a database of student essays (written in English) began in 2014 and the groundwork for an online system to detect plagiarism in the essays collected in that database began the following year. As of the end of September 2016, the database contained approximately 5,000 student essays. The system allows teachers and administrators in the English Department to check the writing of students who have submitted written work against a collection of previously submitted ones using word strings in order to determine cases of plagiarism. The system also facilitates Google searches for potentially plagiarized content using the same word strings.

This project has been part of a comprehensive approach to student plagiarism that has included creating materials to instruct students in the skills of properly citing sources, paraphrasing, summarizing, and quoting; devising a plagiarism policy that may serve as a model for other departments of the university; and making available video tutorials that both warn of the disadvantages and dangers of plagiarism while constructively informing viewers of how it can be avoided.

2. Background

Surveys of attitudes toward, and experiences with, plagiarism were created for the students and teachers in the English Department's Integrated English program. These surveys were administered from the end of 2013 to the beginning of 2014. Results confirmed teachers' perceptions that plagiarism is not practiced by the majority of students but it is a significant problem that must be addressed in a systematic way to ensure fairness and academic integ-

ity. This paper will focus on what the IE Program has done to prevent and combat plagiarism, what other institutions around the world do in that regard that have proven to be helpful, and how we came to develop the teaching materials and online plagiarism detection system that the funding for this project allowed.

Some of the key results of the survey of nearly 400 English Department students (55% of whom were sophomores; 40% freshmen; and the remainder third and fourth year students who were retaking courses) are summarized in figures 1, 2 and 3. As we can see in Figure 1, students report higher instances of plagiarism in the work of classmates than in their own. Since nearly three-quarters of the students skipped the question, it may be the case that approximately 25% either have themselves plagiarized or know of classmates who have. There may have been some students who did not wish to admit to having plagiarized due to the stigma attached to it, despite it being an anonymous survey.

The most common reason for plagiarizing was reported to be "insufficient time to do one's own work" (Figure 2) and the "inability to properly understand the task" (20.23% and 15.27% of the students who plagiarized, respectively). Since it was more common for students to point out the plagiarism of classmates rather than admit to doing it themselves, perhaps it is not surprising that the majority (approximately 60% of the students) hoped for the creation of a plagiarism policy (Figure 3).

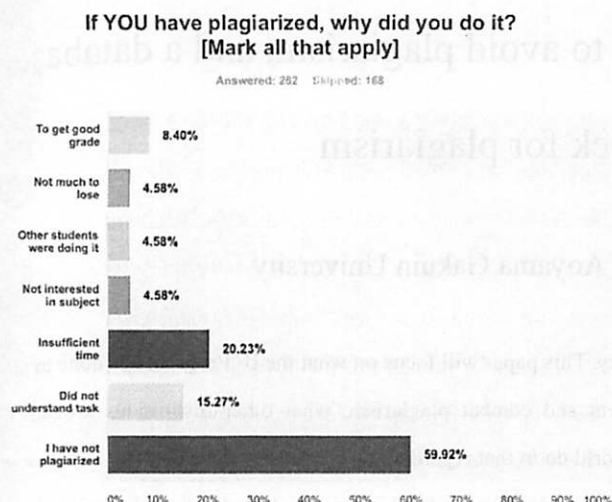


Figure 1: Self-reported rates of plagiarism in various types of assignments among students in the English Department.

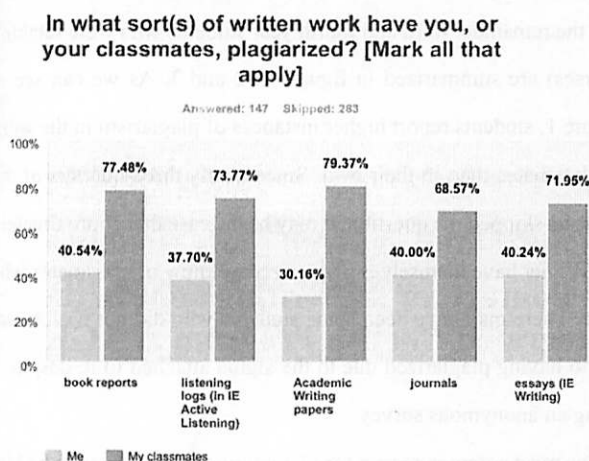


Figure 2: Reasons students who plagiarized gave for doing so.

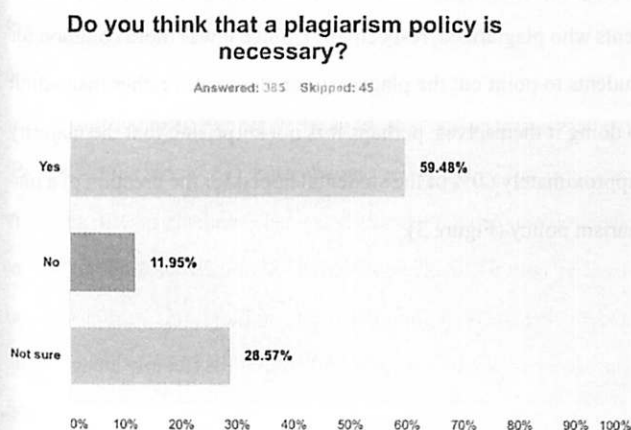


Figure 3: Overwhelming support among students for the institution of a policy concerning plagiarism.

Plagiarism is a complicated problem that can involve educational, legal, career, financial, moral, and emotional challenges. It is a frequent topic among teachers in teachers' rooms around the world and strategies for dealing with it are featured at faculty development sessions and at academic conferences (Plagiarismadvice.org, 2013). Deciding whether a student's work has been plagiarized can be agonizing, especially when the decision is based on an "educated guess" or circumstantial evidence but without iron-clad proof.

Although it is generally acknowledged that a combination of a carrot and stick approach to plagiarism is necessary, it is often the stick that is emphasized in plagiarism statements and policies at universities. Before embarking on writing their master's thesis, Waseda University students in the Graduate School of Economics must sign a document noting that they understand the definition of plagiarism, as stated in their "Study Guide for Students" and that they are aware of the dire penalties associated with plagiarism, which include a suspension from the university for an indefinite period, the loss of credit for their MA thesis, and the invalidation of the most recent semester's credit (Graduate School of Economics, Waseda University).

Waseda University's policy may be perfectly acceptable for graduate students, who presumably should know better by that point in their academic careers, but policies and warnings can only go so far if the causes for the transgressions are not clearly understood and if avoidance through education and motivation has not been fostered. Perhaps, as den Ouden and van Wijk (2011) suggest, the discussion concerning plagiarism should change from "catch-and-punish to teach-and-prevent" (p. 197).

Eisner and Vicinus (2008) note that many instances of plagiarism are due a teacher's failure to properly articulate the parameters of an assignment. If writing tasks are left too open-ended or if only general reflections are solicited, without the inclusion of personal observations or experiences, plagiarism is facilitated. By asking students to use higher order thinking skills (e.g., analysis, synthesis, evaluation), plagiarism is not made impossible, but it becomes a greater challenge—so challenging that producing original work may be the less strenuous option.

3. Definition of Plagiarism

A dictionary definition of plagiarism characterizes it as "the practice of taking someone else's work or ideas and passing them off as one's own" (Oxford Dictionaries Online, N.D.) is not inaccurate but it neglects the very important questions of "what," "how," "why," "in what circumstances," and "to what consequences."

Although the modern concept of plagiarism as being an unethical violation of an author's rights is only a few hundred years old, the idea that copying someone else's work is inappropriate can be traced back to the Roman poet, Martial. When another writer tried to present Martial's poems as his own, Martial lamented that his verse had been kidnapped and labeled the literary thief as a "plagiarist," or kidnapper (Lynch, 2002). The concepts of copyright, authorship, and intellectual property are relatively new even in Western countries and they came about more for economic rather than ethical reasons (Baron, 2000).

4. Varying attitudes toward plagiarism

Second language learners may face greater challenges when paraphrasing and attempting to synthesize multiple sources to create an original work of their own, but the widespread use of costly plagiarism detection systems--such as Turnitin, iThenticate (iParadigm, 2013) and Ephorus (Ephorus International, N.D.)--suggests that the problem is, at least, as serious with those writing in their first languages. Learners in Western countries may be more likely to be exposed, from earlier ages, to expectations that work submitted in school should be "in one's own words" and that the ideas of others should be properly acknowledged.

Despite some instruction on how to cite sources, regulations concerning plagiarism tend to be less strict in Japan and other Asian nations than they are in the West. Tsintzoglou, 2011, in a study on Japanese postgraduate students studying in Australia, showed that they were able to adapt to more severe regulations and the greater emphasis on written work without major problems.

5. Reasons for plagiarizing

Den Ouden & van Wijk (2008) use an aquatic metaphor when discussing the issue of plagiarism in higher education by noting that

the stream of written work becomes "polluted" when a minority of papers submitted are plagiarized, "recycled," or acquired from Internet "paper mills" (Coastal Carolina University, 2009). Students commit plagiarism for a plethora of reasons and through numerous methods. Sometimes it is due to ignorance of what constitutes plagiarism; in extreme cases, the notion that anything on the Internet is fair game for unfettered use. Some cases can be traced to "short cuts" motivated by time pressure or overcommitments. One of the temptations students might have is to self-plagiarize by repurposing work that they previously submitted for a different course by making only cosmetic changes to it. Contextual factors can be involved when students come to view "patch writing," interspersed with some original commentary, as *the norm*, despite a vague awareness of rules to the contrary. At times, plagiarism may simply be the result of careless note taking, when what was paraphrased and what remain direct quotes are confused. Insecurity, and shame of one's own writing, may lead some students to view plagiarism as the "safer" option. Finally, there are students who have an affinity for breaking rules as a kind of thrill seeking.

6. Context for the writing done in the Integrated English Program at AGU

The English Department started an intensive English program, called Integrated English (IE), in 1992 for freshmen and sophomores. Students entering the English Department are placed at an appropriate level based on their performance in the TOEFL ITP. There are three main courses in the program divided into three levels: IE Core, IE Writing, and IE Active Listening. Students in the program take IE Core (180 min), IE Writing (90 min), and IE Listening (90 min) weekly for one semester (15 weeks) at each level. After finishing level 3 (the highest level), students take IE seminars taught by native speakers of English on various academic topics related to linguistics, communication, or literature. In their second year, students take courses in Academic Skills (mainly an academic listening and note taking course) and Academic Writing (See Strong, 2007 for a more detailed description of the program).

Various writing assignments and tasks gradually inculcate increasingly sophisticated writing and critical thinking skills. Students begin to formally learn about MLA style at the IE III level.

the rationale being that they should become comfortable composing paragraphs and short essays based on *their own* ideas before they are asked to cite sources and use academic conventions when referring to content that supports their theses. Students without the ability to compose comprehensible sentences or coherent paragraphs and essays cannot be expected to paraphrase source material and contextualize quotes properly.

In house publications are used to introduce students to library research, ways of selecting sources judiciously (considering authority, currency, relevancy, bias, etc.), using quotations, composing paraphrases, and applying MLA style when inserting in-text citations and creating bibliographies (Strong, 2012; Strong, 2013a).

Writing assignments include informal journal writing, book reports, and response papers on news stories in the Core class; listening logs in IE Active Listening; paragraphs and essays of various types (classification, comparison & contrast, persuasive, analytic) in IE Writing; note taking in Academic Skills; and a 1500 word thesis in Academic Writing.

7. Rationale for having our students write as much as they do

Before we consider plagiarism, it is important to look at why students are asked to write in the first place and the functions that it serves for them, for the teacher, for the institution, for the wider community, and for future careers that students may engage in. Writing is emphasized in the IE Program for some of the same reasons put forward by language programs, and by departments focusing on humanities, around the world. It is so that students...

are able to activate the language they are exposed to through reading and listening.

- * can improve their ability to express themselves.
- * will have a more active attitude toward learning.
- * will be ready to face the demands of the international workplace.

learn to critically assess what they are reading and demonstrate analytical skills.

will have command over different registers of written English (casual, academic, politely informal, etc.)

are prepared for the writing demands of universities abroad, where they may study either as exchange students as undergraduates, or as graduate students.

will not be tempted to plagiarize in the future due to a lack of confidence in their ability or ignorance of academic conventions.

8. Current plagiarism policy

In the 2007/8 academic year, In response to more and more teachers in the program reporting plagiarism on book reports, some explicit penalties were introduced along with minor pedagogical tweaks. These are communicated to teachers through the Integrated English Core 2013 Scope and Sequence (Strong, 2013b).

We required students to inform teachers of the book they had chosen for their book report weeks in advance. Teachers were asked to have the students bring the book to class and the teachers should be especially vigilant of plagiarism if a submitted paper reported on a book other than the one previously announced. Students were also now required--in addition to analyzing the book in respect to literary concepts, such as theme, irony, conflict, climax, symbols, etc.--to answer some specific comprehension questions to make sure that they read the book.

Teachers were asked to "warn students that plagiarism of an assignment [would] lead to a fail on that assignment (no rewrites). If they are caught with a second plagiarism, they [would] fail the entire IE Core class (Strong, 2013b). It was also highly recommended that instructors demonstrate to their students how easy it is to detect plagiarism through Google searches and inappropriate or inconsistent word choice.

The fact that we are now seeing plagiarism in types of writing assignments (e.g., journals) that we had not previously seen demonstrated to us that the further refining of our approach to plagiarism would be necessary. Realizing that a plagiarism policy would be a reasonable first step, we made use of the findings of the surveys we conducted on plagiarism in the English Department at Aoyama

Gakuin University to compose a policy customized to our particular students and their unique learning environment. The policy can be seen in its written form at <http://www.aogaku-daku.org/wp-content/uploads/2015/03/PlagiarismPolicy1.pdf> and being recited by IE Program teachers in a video presentation at: <http://www.aogaku-daku.org/2016/09/18/ie-programs-plagiarism-policy-video-with-subtitles/> (see Figure 4 for a screenshot from the video). The plagiarism statement/policy 1) explains the particular writing requirements for students in the Integrated English Program (IEP); 2) offers a succinct definition of plagiarism; 3) gives guidance on when it is necessary to cite sources; 4) tells how avoiding plagiarism shows our good manners and respect for others; 5) warns of the specific consequences of plagiarism for students in the IEP; and 6) shares a long list of benefits for producing original, plagiarism-free work.



Figure 4: A screenshot from a film featuring IE Program teachers announcing the new plagiarism policy.

9. An explanation of the functions of the database of student writing and plagiarism detection system

The database and plagiarism detection system that has been developed as part of this Information Media Center project offers an interface for students to upload their written work anonymously, and a separate interface for teachers and program administrators to check for possible plagiarism among their students. It is a custom web-based cloud software solution, built on top of open source software running on Apache server architecture, with extensions written in PHP on a framework.

After the completion of the first phase of the project, students

were able to upload their work, and teachers could download those documents. At that stage, the system was used more as a deterrent against plagiarizing than as a way to identify definitive cases of plagiarism. In Phase 2, string searches within the database could be accomplished and Google searches (using the identical strings) were facilitated, as can be seen in the screenshots shown in Figures 5-10.



Figure 5: Login page for both students and teachers.

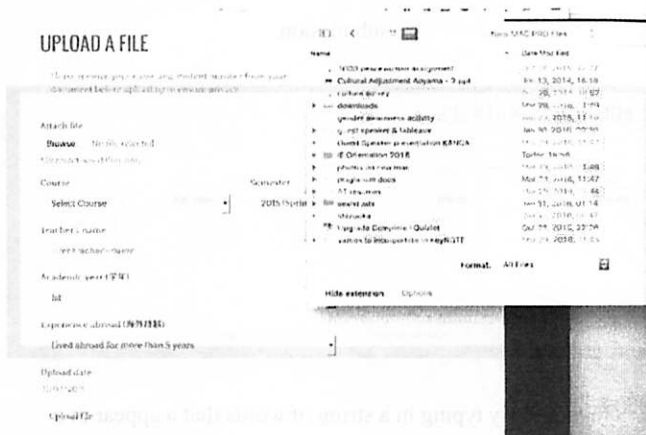


Figure 6: Page that students use for uploading their written work to the database of students writing. Students browse for their essay on their local disk and upload it.

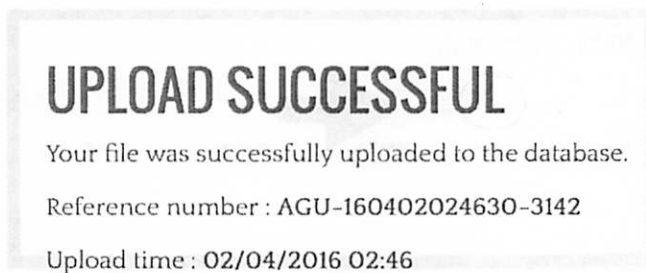


Figure 7: Students receive a reassuring message that their essay was uploaded safely along with a unique identifier (i.e., code) that they must submit to their teacher to prove that their essay was uploaded and to allow the teacher to connect that particular

essay to the student, making it unnecessary for students to include any personal identifying information (such as name or student number) on their essay document.



Figure 8: Teachers may then go into the system and download essays that their students have submitted, searching for them using the unique reference number issued to students at the time of submission.

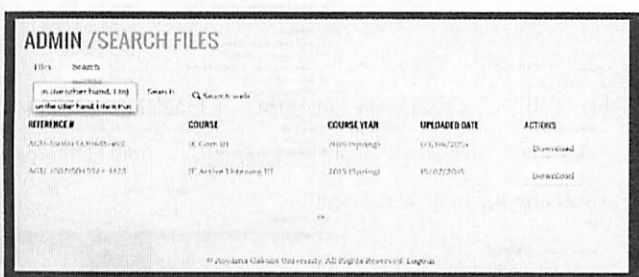


Figure 9: By typing in a string of words that appear in a particular essay, other documents with an identical string of words can be detected and compared for overall degree of commonality in order to determine if a case for plagiarism can be made.

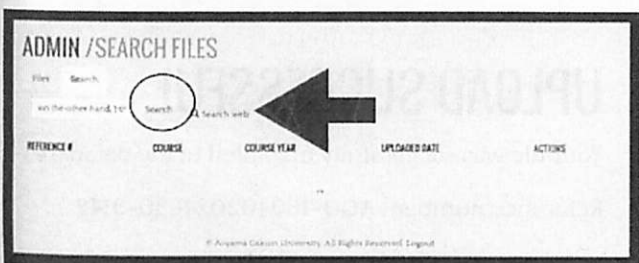


Figure 10: By typing in a string of words that appear in a particular essay, other documents contained in the database exhibiting an identical string of words can be detected and the essays can be compared for their overall degree of commonality in order to determine if a case for plagiarism can be made.

10. Future Directions

Phases 3 and 4 of the project will involve the more complex task of creating custom algorithms to extend the plagiarism detection system by offering these enhanced functions:

- * Improved security through unique logins for teachers
- * The ability for teachers to view essays by individual class
- * Automatic detection and flagging of suspected plagiarized essays and the display of the percentage of overlap, and
- * Side-by-side presentation of suspected plagiarized item with possible source (i.e., common elements) highlighted.

References

- Avoiding Plagiarism. (n.d.). University of Leicester's Learning Development. Retrieved September 11, 2013, from www2.le.ac.uk/offices/ssds/sd/ld/resources/study/plagiarism-tutorial
- Baron, N. (2000). Alphabet to email: How written English evolved and where it's heading. London: Routledge.
- Coastal Carolina University. (2009). Cheating 101: Internet Paper Mills. Kimbel Library: Presentations. Retrieved September 9, 2013, from <http://www.coastal.edu/library/presentations/mills2.html>
- Den Ouden, H., & Van Wijk, C. (2011). Plagiarism: Punish or Prevent? Some Experiences With Academic Copycatting in the Netherlands. *Business Communication Quarterly*, 74(2), 196–200.
- Eisner, C., & Vicinus, M. (Eds.). (2008). Originality, imitation, and plagiarism: Teaching writing in the digital age. Ann Arbor: University of Michigan Press.
- Emma, L. (n.d.). Consequences of Plagiarism & Cheating. Retrieved September 5, 2013, from <http://everydaylife.global-post.com/consequences-plagiarism-cheating-6118.html>
- Ephorus Home Page. (n.d.). Ephorus International. Retrieved September 10, 2013, from <https://www.ephorus.com/>
- Evaluation Wizard. (2012). 21st Century Information Fluency. Retrieved September 11, 2013, from <http://21cif.com/tools/evaluate/>

- Fox, K. (2010). *Plagiarism! Plagiarism!: 25 Fun Games and Activities to Teach Documenting and Sourcing Skills to Students*. Madison (Wisconsin): Upstart Books.
- Harris, R. (2012). *Anti-Plagiarism Strategies for Research Papers*. VirtualSalt. Retrieved September 7, 2013, from <http://www.virtualsalt.com/antiplag.htm>
- Hyland, T. A. (2009). Drawing a Line in the Sand: Identifying the Borderzone Between Self and Other in EL1 and EL2 Citation Practices. *Assessing Writing*, 14(1), 62–74.
- iParadigms. (2013a). Turnitin Home Page. Retrieved September 15, 2013, from <http://turnitin.com/>
- iParadigms. (2013b). iThenticate. Retrieved September 23, 2013, from <http://www.ithenticate.com>
- Lynch, J. (2002). The Perfectly Acceptable Practice of Literary Theft: Plagiarism, Copyright, and the Eighteenth Century. *The Journal of the Colonial Williamsburg Foundation*, 24(4), pp. 51–54.
- Place, E., Kendall, M., Hiom, D., Booth, H., Ayres, P., Manuel, A., & Smith, P. (2006). Internet Detective: Wise up to the Web. Intute Virtual Training Suite. Retrieved September 11, 2013, from <http://www.vtstutorials.ac.uk/detective/>
- Plagiarismadvice.org. (2013). 6th International Integrity & Plagiarism Conference. Retrieved September 22, 2013, from <http://6ipc.eventbrite.co.uk/>
- Strong, G. (2007). Curriculum Design: Furniture for a college EFL program. In M. Carroll (Ed.), *Developing a New Curriculum for Adult Learners* (pp. 153–176). Alexandria (Virginia): TESOL.
- Strong, G. (2012). *Academic Writing Students' Booklet*. Tokyo: Aoyama Gakuin University. Retrieved from <http://www.cl.aoyama.ac.jp/~dias/docs/2012-AW-Student-Guide.doc>
- Strong, G. (2013a). *Integrated English Core and IE Writing*. Tokyo: Aoyama Gakuin University. Retrieved from <http://www.cl.aoyama.ac.jp/~dias/docs/2013-NEWIECore and Writing.docx>
- Strong, G. (2013b). *Integrated English Core 2013 Scope and Sequence*. Tokyo: Aoyama Gakuin University. Retrieved from http://www.cl.aoyama.ac.jp/~dias/docs/2013-NEW-REV--Scope_AND_Sequence.doc
- Sunday Morning (CBS News). (2012). The problem with plagiarism. Retrieved September 19, 2013, from <http://www.cbsnews.com/video/watch/?id=50133558n>
- Teach Information Literacy & Critical Thinking. (2013). UCLA College Library Librarians. Retrieved September 12, 2013, from <https://sites.google.com/site/teachinfolit/find-evaluate-web-sites-blogs-wikis-more/hoax-scholarly-research-personal-opinion-you-decide>
- Tsintzoglou, G. (2011). Japanese postgraduate students' perceptions of and attitudes towards plagiarism. The University of Melbourne.
- Understanding Plagiarism. (2004). Indiana University, School of Education Web Site. Retrieved September 11, 2013, from <https://www.indiana.edu/~tedfrick/plagiarism/>
- Waseda University Graduate School of Economics. (n.d.). Note of Recognition of the Definition of Plagiarism. Retrieved from http://www.waseda-pse.jp/gsc/en/common/pdf/20121101_hyosetsuteigisyo_eng.pdf
- Weber-Wulff, D. (2008). On the Utility of Plagiarism Detection Software. Retrieved September 7, 2013, from <http://plagiat.htw-berlin.de/softwareutility/>