Publishing in a World with "Plagiarism" (i.e., Similarity) Detection Software.

Abstract

The literature on academic integrity and plagiarism is often described as a war, battle, or race between two opposing sides. Consider, for example, the following quote from an article on automated paraphrasing tools: "…there is a technological arms-race occurring between the development of tools and techniques which facilitate violations of the principles of educational integrity, including text-based plagiarism, and methods for identifying such behaviors" (Roe & Perkins, 2022).

And, as with any battle, there is a good chance that some non-combatants may get caught in the crossfire. Academic researchers may be one of those non-combatants. Many editors are now using similarity check software (e.g., CrossRef, Turnitin) during their initial review of a submitted manuscript; and as a result, manuscripts submitted for review by journals are being flagged more frequently for having a high degree of similarity with previously published articles. While the producers of this type of software make a point of stating that they are not plagiarism detection software, they do suggest that their software is a useful tool for identifying potential plagiarism. Thus, it would not be unusual for editors dealing with a high volume of initial reviews to place too much weight on the score that a manuscript receives from the software. This is especially an issue for research-oriented manuscripts, because the software producers note that well-researched manuscripts (e.g., a manuscript that contains a large number of relevant citations) will earn a high score. In fact, we may want a research-oriented manuscript to earn a high score. For this reason, editors need to think more critically about what a high score represents and what steps to take to determine if a high score represents a strength of the manuscript or is truly a sign of plagiarism.

During our presentation we will share a manuscript that was rejected for having a high similarity score, along with the editor's comments. We will identify some of the characteristics of the manuscript that created the high similarity score and discuss the efforts that we tested to modify the manuscript so that it would earn a lower score. In most cases, we think that the changes (e.g., revising wording in statistical analysis section) decreased the quality of the manuscript. We will end the presentation with a discussion about implications this software has for editors and researchers alike.