

Plagiarism and self-plagiarism: What every author should know

Miguel Roig

Department of Psychology, St. John's University, Staten Island, New York, USA

Corresponding author: roigm@stjohns.edu

Abstract

The scientific community is greatly concerned about the problem of plagiarism and self-plagiarism. In this paper I explore these two transgressions and their various manifestations with a focus on the challenges faced by authors with limited English proficiency.

Keywords: plagiarism; self-plagiarism; limited English proficiency

Received: July 1, 2010

Accepted: September 10, 2010

Introduction

Evidence indicates that plagiarism amongst biomedical students is fairly common (1-3). Because the offenses in question usually involve academic assignments, they are typically classified as instances of academic dishonesty. Such transgressions can result in negative consequences for the student and these can range from failure for the assignment to expulsion from the university. When plagiarism occurs in the context of conducting scientific research, whether perpetrated by students or by professionals, it rises to the level of scientific misconduct; a much more serious crime.

Regrettably, a general consensus is now emerging that plagiarism in the biomedical sciences has become a matter of great concern. Consider the evidence, when searching the PubMed database for articles on plagiarism (4), the database yields over 700 entries (as of this writing) with more than half of them representing articles that were published within the last decade. Also, journals are increasingly expanding their instructions to authors to include guidelines on plagiarism and related matters of authorship. Yet, perhaps the most alarming development has been the availability of text similarity software, such as eTBLAST, that allows users to

search for plagiarism in journal articles (5). Given these developments, it is not surprising that a recently published survey shows plagiarism as one of the areas of greatest concern for biomedical journal editors (6).

The causes underlying many cases of plagiarism are believed to be the same as those associated with the other two major forms of scientific misconduct, fabrication and falsification. For example, one major factor believed to operate is the pressure to publish. The reality is that for many working scientists, the number of published papers authored continues to be one of the primary means by which research productivity is measured. Moreover, the quality of a publication is another important factor that comes into play, for the most desirable outcome is for papers to appear in the so-called high-impact journals. Of course, carrying out scientific research can be very rewarding intrinsically and the joy we experience when we are engaged in this noble process is probably the very reason why many of us chose science as a career. However, as we all know, good science requires a lot of patience, hard work, and a good dose of creative, methodological skill. In addition, scientific

research has become very costly in terms of human and laboratory resources. Our tenacity and dedication will usually pay off, as when we are able to obtain data that verifies our hypotheses. But as every scientist knows, such a happy ending does not always occur. For example, what at first might look like a promising avenue of investigation can sometimes end up being a dead-end. In a worst case scenario, months of toiling in the laboratory may only yield a limited payout as when results turn out marginal or null and, therefore, not likely to be publishable. Or perhaps a subtle mistake early in the experiment can render as useless months of otherwise meticulous laboratory work. These are some of the many scenarios that are thought to lead otherwise well-meaning scientists to tamper with their data.

Because plagiarism and self-plagiarism are thought to be far more common than fabrication and falsification, it is important to explore these transgressions in some detail. The reader should note that these offenses can sometimes have legal implications, as when they violate copyright law. However, because these cases rarely, if ever, reach the legal stage when they involve scholarly journals, I will confine my treatment of these malpractices within the ethical domain rather than within the legal one. My hope is that, by raising the readers' awareness of these offenses, their occurrence can be prevented.

Plagiarism

Writing journal articles is seldom an easy task and many of us do not exactly enjoy this part of the scientific process. To make matters worse, we often operate with the expectation that our manuscript will be returned with a myriad of criticisms and suggestions for improvement that are sometimes viewed by us as arbitrary and capricious. Although this feedback almost always results in an improved product, I suspect that most authors dread this aspect of the process and few of them genuinely welcome such efforts. In the end, however, most of us recognize that the peer review system is an integral part of the cycle of science.

Good writing is seldom easy to produce and effective scientific prose can take time and much mental effort to generate even for experienced authors. Thus, the temptation to look for short-cuts can arise particularly if the author is experiencing some form of 'writers' block', a temporary inability to become inspired and produce new work. In these situations, the urge to 'borrow' others' well-crafted prose may be irresistible. But, one might ask, what is the harm in such borrowing? After all, taking a couple of lines of text does not, in any way, affect the integrity of the data and it is the latter that is most important (7). Besides as an ethical offense in the sciences, plagiarism of text is arguably far less serious than plagiarism of ideas or plagiarism of data (8). Moreover, since there is no universally agreed-upon operational definition of plagiarism in terms of how many consecutive words can be copied without attribution, who is to say that it is wrong to appropriate a well-written sentence or two that elegantly conveys a very complex process or phenomenon? Other considerations seem to even favor such minor 'borrowing'. For example, when describing a highly technical methodology and/or procedure commonly used by our peers, there is some risk that even a small change in the wording could result in subtle misinterpretations of the methods or procedure and that possibility is highly undesirable (9). Of course, the latter rationale is a poor excuse for the copy-pasting of large segments of methodology sections. Besides, in the quest for conciseness, these sections sometimes lack some important details and, therefore, can often benefit from rewriting for purposes of enhancing their clarity (10). Unfortunately, there are those, whose writing style is such that they take a liberal approach to using others' text as their own (11). But, in the current climate of responsible research conduct, such writing practices now run a greater risk of being noticed and, at best, they will be judged with suspicion, for they certainly do not represent high standards of scholarship.

It is totally understandable when the main reason given for using others' text is lack of language/writing proficiency (12). However, as much as we can empathize with such authors, the scientific community could not function properly with different

scholarship criteria depending on one's level of language proficiency (10). The reality of the situation is that English has become the *lingua franca* of science and most, if not all, of the high impact factor journals are published in English. Even some of the journals published in non English-speaking nations are published in English, i.e., *Biochemia Medica*, and the expectation is for scientists from these nations to also publish in English. This situation presents a unique challenge for the Limited English Proficiency (LEP) author, but even some of these authors recognize that it is a challenge that must be met (13). English is not an easy language to learn, especially for those whose native language is based on a different alphabet system. Moreover, while good skills in English are necessary for writing journal articles, they are not sufficient to do the job. To write effective scientific prose, not only do we need to be proficient in the language, we also need to have a thorough grasp of the technical language and the unique expressions and phraseology associated with the particular knowledge domain in question. In other words, we need to be able to understand what we are reading and also to convey that information using our own words and domain-consistent expressions; our own 'voice'. In fact, evidence that I have collected in the past suggests that text readability is a strong predictor of misappropriation not only by students (14) but also by professors (15). Novice researchers and especially LEP authors will often encounter these types of reading/writing difficulties when dealing with unfamiliar technical literature in their disciplines. Therefore, I strongly believe that these are the very factors that are behind a significant amount of plagiarism.

Does 'borrowing' a few sentences here and there (i.e., patchwriting) rise to the level of plagiarism? I suppose that it depends on the circumstances, the number of sentences that have been misappropriated and on who is doing the judging. However, the fact remains that passing as one's own the work of others, even if it is a small amount, is consistent with any definition of plagiarism. In addition, such practices are now more likely to be discovered given the availability of software programs designed to detect plagiarism. For example, consider

the recent case in which a paper was retracted from a journal because merely two paragraphs from its introduction were found to be identical to paragraphs appearing in an earlier published paper by a different author (16). The message is clear: Using textual material without proper attribution is plagiarism, even when it is done in relatively small amounts.

Self-plagiarism

Whereas plagiarism involves the presentation of others' ideas, text, data, images, etc., as the products of our own creation, self-plagiarism, occurs when we decide to reuse in whole or in part our own previously disseminated ideas, text, data, etc without any indication of their prior dissemination. Perhaps the most commonly-known form of self-plagiarism is duplicate publication, but other forms exist and include redundant publication, augmented publication, also known as meat extender, and segmented publication, also known as salami, piecemeal, or fragmented publication. The key feature in all forms of self-plagiarism is the presence of significant overlap between publications and, most importantly, the absence of a clear indication as to the relationship between the various duplicates or related papers. Because of the latter, the word 'covert' should always be added to these designations (e.g., covert duplicate publication, covert redundant publication, etc.). As with traditional forms of plagiarism, a very likely cause of much self-plagiarism appears to be authors' desire to add publications to their vita (17).

In a typical duplicate publication, authors of a previously published paper submit roughly the same manuscript to a different journal. The second submission may have a slightly different title, a different order of authorship, perhaps minor changes to the text of the manuscript, but the data and statistical analyses are largely the same. These instances of duplication are typically easy to spot because the identical text, formatting, data tables, etc., are usually recognized by the astute reader who is familiar with that specific area of research. A more harmful version of duplicate publication occurs when the authors make an effort to conceal the

fact that the same data are being republished more than once. In these cases the perpetrator makes a concerted effort to make significant textual changes to various components of the paper, such as the literature review, discussion, etc., and they may do so by, for example, adding and/or deleting certain references. Furthermore, the formatting of tables of data and of graphs may also be changed, thus giving the appearance of a different set of data and a distinct paper. Again, the key component of this malpractice is that the new paper makes no reference to the previous publication, or if it cites the previous paper, it does so in such an ambiguous manner that the reader fails to recognize the exact relationship between the two papers, thus the term covert duplicate.

There can be various other permutations of this basic approach and von Elm and his colleagues have described a number of them (18). In one version, for example, authors of a previously published paper may reuse its data and carry out a different set of statistical analyses. The results of these analyses are then included in a paper whose title, abstract and portions of the introduction and discussion may now be somewhat different in the context of these new analyses. In another version, data from two or more previously published papers are presented together as new with perhaps additional statistical analyses included. In instances of augmented publication, or meat extender as this type of redundancy is sometimes called, authors simply add additional observations or data points to a previously published data set. They then reanalyze the augmented data set, and publish a paper based on the new results. Again, it is important to emphasize that such practices may be acceptable if the author provides the editor with a defensible rationale for his actions and makes it clear to the reader that the data are derived, in whole or in part, from a previous publication. However, because most journals only accept original research, such a clarification often renders the paper unsuitable for publication. Again, because publication of the new paper is the primary aim for the unscrupulous author, this fact tends to remain hidden from the editor and the reader.

Segmented or salami publication is a distinct publication practice that may, in theory, contain little if any self-plagiarized text and/or data. However, even in the absence of any text or data reuse, the practice is nevertheless, problematic and actively discouraged in the sciences. A typical case involves a complex experiment/study (i.e., the whole salami) that yields multiple measures or sets of measures from the same study sample. Rather than publishing the results of these various data sets together in a single publication, the investigators analyze and publish each data set separately (i.e., salami slices). In this way the single experiment can yield two or more articles thereby enhancing the investigators' publication list. As in other forms of covert redundancy and covert duplication, this practice is considered unethical if each salami slice (i.e., segmented publication) fails to reveal the fact that its data are derived from the same experiment as data from other related publications that were part of the same salami.

There can be legitimate reasons for the various forms of redundancy. For example, with respect to salami publication, it is not uncommon in longitudinal-type studies, such as the Framingham Heart study (19), for different sets of authors to publish observations from the same longitudinal sample in separate journal articles. This is completely acceptable and even desirable when the interval of time between observations made from the sample spans years. Likewise, for other types of experiments there may be good reasons to report different results arising from a single experiment in two or three different journals as the various observations may be of interest to different audiences. However, authors must always inform readers about the exact origin their data and how their data are related to other published papers. Even duplicate publications may be totally acceptable as when a paper first appears in one language and it is then translated into another language and published in a different journal or edited volume. But, again, the second publication must always provide a clear indication as to its association with the earlier published version.

The major scientific organizations (e.g., Committee on Publication Ethics, World Association of Medi-

cal Editors) and even individual journals offer relevant guidelines to avert instances of self-plagiarism. For example, the Uniform Requirements for Manuscripts Submitted to Biomedical Journals" (20) published by The International Committee of Medical Journal Editors calls on authors to inform the editor of the journal, upon submission of a manuscript, to reveal other related published papers or manuscripts that have been prepared for other journals.

Obviously the primary issue in self-plagiarism (i.e., duplicate, redundant publication, and augmented publication) concerns the covert reuse of already published data that are being portrayed as new data. In the case of salami publication the main concern is the presentation of data sets that are portrayed as having been independently derived when in fact they come from a study from which other related data were collected. The problem with such misleading portrayals of data is that they are likely to mislead others by overestimating, or depending on the type of problem being addressed, underestimating a particular effect or process. For example, let's assume that there exist various covert duplicates that show a certain drug to be highly effective as a cure for a disease. Someone conducting a meta-analysis on the efficacy of the drug may be unaware that some of the studies found are actually cleverly disguised covert duplicates of existing ones. The inclusion of these duplicates results in an inflated effect size, which in turn distorts researchers' understanding of the true effectiveness of the drug (21).

One last form of self-plagiarism that must be discussed, and one that I believe to be most strongly related to language proficiency is what some refer to as same-authored text recycling. A typical instance of this practice occurs when authors reuse large portions of text that they have already published in one or more journal articles and these are then reused in a new publication (9,22). For the native speaker/writer, the practice represents, at best, a case of intellectual laziness (23) or poor scholarly etiquette and is certainly discouraged by some journals (24). Text recycling, when practiced out of necessity by LEP authors, certainly does not merit such negative characterizations. However, it is still deemed as a problematic practice.

Why should we be discouraged from reusing textual material that we ourselves have produced? Here are some reasons. I believe that there is an underlying assumption on the part of the author who is engaged in these practices, that the previously written material is so well crafted and clear that it cannot benefit from improvement (10,25). In my experience as a reader of primary literature and as a journal reviewer, I often find that assumption to be totally unwarranted. In addition, merely relying on copy-pasting to create a methodology section runs the risk of failing to include or exclude crucial details unique to the new experiment being described. There is at least one editor that cautions potential authors against the mere recycling of previously published methods sections without modification (26) and already one study has uncovered evidence of important lapses when using copy-pasting techniques with medical records (27). Thus, relying on mere copying and pasting of text can be highly problematic when used in scientific articles. Equally important perhaps, is the fact that text recycling does not constitute scholarly excellence, for it violates a basic assumption of the implicit reader-writer contract. Accordingly, the reader operates under the assumption that 1) the author/s is the individual who produced the work, 2) any text, ideas, etc., that are taken from other available sources, even if produced by the same author, are identified with standard scholarly conventions, such as citations and quotations, and 3) that the ideas, data, etc. presented are accurate (28).

In sum, plagiarism and self-plagiarism can manifest themselves in a variety of forms. Depending on the circumstances, these transgressions can merit labels that range from poor or sloppy scholarship to scientific misconduct. Some LEP authors may be particularly vulnerable to excessive 'borrowing' from others' work as well as from their own previously published papers. While their situation is totally understandable they should keep in mind that most of us in the scientific community regard science as highest form of scholarship. As such, we expect nothing but the highest standards of practice from those who are given the privilege of engaging in this most noble of activities.

References

- Ryan G, Bonanno H, Krass I, Scouller K, Smith L. Undergraduate and postgraduate pharmacy students' perceptions of plagiarism and academic honesty. *Am J Pharm Educ* 2009;73:105.
- Rennie SC, Crosby JR. Are "tomorrow's doctors" honest? Questionnaire study exploring medical students' attitudes and reported behaviour on academic misconduct. *BMJ* 2001;322:274-5.
- Billić-Zulle L, Frković V, Turk T, Petrovečki M. Prevalence of plagiarism among Medical students. *Croat Med J* 2005;46:126-31.
- Aronson JK. Plagiarism - please don't copy. *Br J Clin Pharmacol* 2007;64:403-5.
- Errami M, Garner H. A tale of two citations. *Nature* 2008;451:397-99.
- Wager E, Fiack S, Graf C, Robinson A, Rowlands I. *J Med Ethics* 2009;35:348-53.
- Yilmaz I. Plagiarism? No, we're just borrowing better English. *Nature* 2007;449:658.
- Bouville M. Plagiarism: words and ideas. *Sci Eng Ethics* 2008;14:311-22.
- Roig M. Re-using text from one's own previously published papers: an exploratory study of potential self-plagiarism. *Psychol Rep* 2005;97:43-9.
- Roig M. Plagiarism: Consider the context (letter to the editor). *Science* 2009;325:813-4.
- Julliard K. Perceptions of plagiarism in the use of other author's language. *Fam Med* 1994;26:356-60.
- Vasconcelos S, Leta J, Costa L, Pinto A, Sorenson MM. Discussing plagiarism in Latin American science. *EMBO reports* 2009;10:677-82.
- Affi A. Plagiarism is not fair play. *Lancet* 2007;369:1428.
- Roig M. When college students' attempts at paraphrasing become instances of potential plagiarism. *Psychol Rep* 1999;84:973-82.
- Roig M. Plagiarism and paraphrasing criteria of college and university professors. *Ethics Behav* 2001;11:307-23.
- Science Insider. From the Science Policy Blog. *Science* 2009;325:527.
- Yank V, Barnes D. Consensus and contention regarding redundant publications in clinical research: cross-sectional survey of editors and authors. *J Med Ethics* 2003;29:109-14.
- von Elm E, Pogia G, Walder B, Tramér M R. Different patterns of duplicate publication: An analysis of articles used in systematic reviews. *JAMA* 2004;291:974-80.
- Framingham Heart Study. Available at: <http://www.framinghamheartstudy.org/>. Accessed April 2nd 2010.
- Redundant publication. Uniform Requirements For Manuscripts Submitted To Biomedical Journals: Writing and Editing For Biomedical Publication. Updated October 2007. Available at: <http://www.icmje.org/>. Accessed March 6th 2010.
- Tramèr M, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *BMJ* 1997;315:635-40.
- Bretag T, Carapiet S. A Preliminary Study to Identify the Extent of Self-Plagiarism in Australian Academic Research [Electronic version]. *Plagiary* 2007;2:92-103. Accessed January 5th, 2010, from <http://hdl.handle.net/2027/spo.5240451.0002.010>.
- Editorial. Self-plagiarism: unintentional, harmless or fraud? *Lancet* 2009;374:664.
- Griffin GC. Don't plagiarize - even yourself! *Postgrad Med* 1991;89:15-6.
- Roig M. The debate on self-plagiarism: inquisitional science or high standards of scholarship. *J Cogn Behav Psychother* 2008;8:245-58.
- Biros MH. Advice to Authors: Getting Published in Academic Emergency Medicine. Available at: <http://www.saem.org/inform/aempub.htm>. Accessed March 6th 2003.
- Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic records trustworthy? Observations on copying, pasting and duplication. *AMIA Annu Symp Proc* 2003:269-73.
- Roig M. Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing (U.S. Department of Health and Human Services, Office of Research Integrity). Available at: <http://ori.hhs.gov/education/products/plagiarism/>. Accessed January 5th 2010.

Plagiranje i samoplaganje: Što bi trebao znati svaki autor

Sažetak

Znanstvena zajednica je vrlo zabrinuta zbog problema plagiranja i samoplaganja. U ovom članku istražujem te dvije transgresije i njihove brojne manifestacije sa žarištem na izazovima s kojima su suočeni autori s ograničenim vještinama uporabe engleskog jezika.

Ključne riječi: plagiranje; samoplaganje; ograničena vještina uporabe engleskog jezika