Article

Plagiarism or Not? Investigation of Turnitin®-Detected Similarity Hits in Biology Laboratory Reports[S]

Tricia Bertram Gallant†
Madeleine Picciotto‡
Goran Bozinovic§
Ella Tour ¶*

From the †Academic Integrity Office, University of California, San Diego, La Jolla, California, 92093, ‡Analytical Writing Program, University of California, San Diego, La Jolla, California, 92093, §Division of Biological Sciences, University of California, San Diego, La Jolla, California, 92093, ¶Section of Cell and Developmental Biology, Division of Biological Sciences, University of California, San Diego, La Jolla, California, 92093

Abstract

In undergraduate biology laboratory courses, laboratory reports can be a useful tool for teaching scientific writing, integration of source material, and information literacy; however, these teaching objectives are at times undermined by students' plagiarism. Laboratory instructors often use similarity-matching software to detect plagiarism in laboratory reports, yet similarity hits detected with such software remain poorly characterized. In the upper division molecular biology laboratory course described here, Turnitin® routinely detected dozens of similarity hits in laboratory reports. To determine whether this abundance of similarity hits was indicative of widespread plagiarism, we analyzed similarity hits detected in 255 laboratory reports written by 135 students. Only a small minority of Turnitin® similarity matches were problematic, but over half of the laboratory reports contained at least one problem with incorporation of scientific sources (e.g., laboratory manual and scientific articles). We identified four common types of such writing problems: patchwriting, technical parroting, copying, and falsification of sources. In 18% of the laboratory reports, we detected an alarmingly superficial use of primary literature. Most of the source incorporation problems did not rise to the level of plagiarism. As a result of this study, we recommend changes in scientific writing instruction and a transition to laboratories providing more authentic research experiences. © 2019 International Union of Biochemistry and Molecular Biology, 47(4):370–379, 2019.

Keywords: *Plagiarism; source incorporation; laboratory reports; undergraduate laboratory courses; similarity-matching software*

Introduction

In undergraduate biology classes, laboratory reports whose structure parallels that of scientific articles have been seen as a standard tool for acquiring writing competency, as they aim to guide students toward critical and evaluative thinking, information literacy, clearly written scientific communication, and an appropriate integration of source material. It has been suggested that the process of writing a laboratory report enables students to make sense of their laboratory experience within the context of scientific inquiry [1].

A traditional undergraduate biology laboratory report includes several standard sections that mirror the structure of a scientific paper: a brief Abstract summarizes the most important findings, Introduction section explains the background of the experiment, Materials and Methods section outlines and details experimental procedures, Results section presents data, and Discussion section elaborates on data analysis. This traditional format was implemented in the upper division molecular biology laboratory classes that were the focus of this study. For the past seven years of teaching these laboratories, similarity matching software, Turnitin®, has been used to detect potential instances of plagiarism. Over the years, instructors have been aware of the large number of similarities detected by Turnitin® in many of the laboratory reports. The majority of these matches linked to laboratory reports by other students, which raised the possibility of our students having access to, and plagiarizing from, these laboratory reports.

Volume 47, Number 4, July/August 2019, Pages 370–379
**To whom correspondence should be addressed. Tel.: (858) 534-0913; Fax: (858) 822-3021. E-mail: etour@ucsd.edu*
[S] Additional Supporting Information may be found in the online version of this article.
Received 20 November 2018; Revised 7 February 2019; Accepted 6 March 2019
DOI 10.1002/bmb.21236
Published online 28 March 2019 in Wiley Online Library (wileyonlinelibrary.com)

A growing incidence of plagiarism among science students plagiarism, a deliberate use of someone else's ideas and language without acknowledgment of their source [2], has been a source of great concern in academia [3–7].

Since the literature on plagiarism in laboratory reports has been limited to surveying students' self-reported behaviors such as copying from other students or falsifying data [8], we decided to conduct an examination of a large number of laboratory reports in order to understand the nature of the matches detected by Turnitin®. Specifically, we asked: do the abundant similarity matches identified by Turnitin® in student laboratory reports indicate plagiarism or misuse of sources? Another possibility is that these matches result from unavoidable repetition of technical terms and protocol details, resulting when hundreds of students write reports on the same experimental procedure. To answer this question, we conducted a detailed analysis of Turnitin®-detected similarity matches in 255 laboratory reports written by 135 students in this course.

Methods

Context of the Study

This study was conducted at a large, research intensive, public university where about 800 students take the upper division molecular biology laboratory course annually. The participants of this study were 135 students, mostly seniors majoring in Biological Sciences (92%), enrolled in three of the laboratory sections during the 2012–2013 academic year. One instructor (G.B.) taught two of the sections (35 and 49 students, respectively), and another instructor (E.T.) taught the third section (51 students).

Students typically wrote three to four laboratory reports, each laboratory report addressing a project that spanned three or more laboratories. Students submitted the laboratory reports were submitted to Turnitin®. The course laboratory manual provided students with the background to the experiments and with detailed step-by-step protocol instructions. Two laboratory reports from each student were examined in this study: Laboratory Report 1, based on the cloning of the *lux* operon from the bioluminescent bacterium *Vibrio fischeri,* and Laboratory Report 3, based on phenotypic observation and qRT-PCR analysis of an RNAi-mediated knockdown of *unc-22* gene expression in *C. elegans* The two instructors used the same guidelines for the laboratory reports. The same experiments had been conducted for at least five (Laboratory Report 3) to eight (Laboratory Report 1) years, so the pool of existing laboratory reports written by former students and submitted to Turnitin® was extensive. Information about academic integrity was provided to students in the syllabus and was posted on the course website (Supporting Information Appendix A).

Turnitin® Reports

Detailed description of the analysis of Turnitin®-generated reports is provided in the Supporting Information, Appendix B. Briefly, we analyzed 255 Turnitin® originality reports (Laboratory Report 1: 130 reports and Laboratory Report 3: 125 reports). All matches were examined by a team of two biology instructors (E.T. and G. B.) with the goal to determine if a match was significant or if the similarity could be attributed to chance alone. Others have deemed the appropriation of five or more consecutive words from a source as constituting an incident of plagiarism [9, 10]. However, in this study, the laboratory reports contained a substantial amount of technical terminology that could not have been expressed differently. Therefore, in considering if a match was significant, we asked, "In how many other ways could the same information be communicated? When the length of the match was relatively short and the information very technical or included commonly used terminology, we considered the match insignificant (Table I). After several joint norming sessions, the two biology instructors made the determination of significance independently and then met and discussed any disparities, until consensus was reached.

The next task was to determine the source text of each match. Turnitin® very reliably identified matches to journal articles and websites, however, the majority of the similarity matches were attributed by Turnitin® to other students' papers and for these matches, the source text was not readily available. Furthermore, we determined that most of such

TABLE I ***Examples of Turnitin® matches that were determined to be insignificant***

Student's text	*Matched source text*
As it is found predominantly in symbiotic relationships with marine organisms, the **symbiosis between *V. fischeri* and the Hawaiian bobtail squid *Euprymna scolopes* has** been carefully studied by scientists for many years.	Interest in the light-organ **symbiosis between *V. fischeri* and the Hawaiian bobtail squid *Euprymna scolopes* has** led several researchers to adopt strain ES114 (4)... Source: journal article, Lyell *et al.*, 2008
Inside the cells, the dsRNA is chopped up with the **enzyme Dicer into double stranded small interfering RNAs (siRNA)** that group together into what is called a RNA-induced silencing (RISC) complex, held together by argonaute proteins.	Once dsRNA enters the cell, it is cleaved by an RNase III – like **enzyme, Dicer, into double stranded small interfering RNAs (siRNA)** 21–23 nucleotides in length that contain 2 nucleotide overhangs on the 3′ ends (9–11). Source: website, http://www.genelink.com/sirna/RNAicustom.asp

Identical text is indicated in bold. Even though both examples contain a sequence of words that is identical to the matched source (12 identical words in a row in the first example, 9 in the second), we determined that it would be very difficult to deliver the same information in other ways and therefore designated these and similar matches as insignificant.

matches were actually matches to the laboratory manual. Table II shows an example from one student's paper, where Turnitin® found matches to 47 different students' laboratory reports, all from our institution. All 47 of these matches had similarities to the laboratory manual text as well, making it by far the most likely source. When this type of source misattribution was encountered, we treated all such matches as one match to the laboratory manual. In the majority of the cases, a consensus was reached on significant matches and their likely sources. Cases where no consensus was reached were excluded from the analysis.

After significant matches were determined, the second coding team (M.P., a writing expert and T.B.G., a plagiarism expert) categorized these matches into types of source material incorporation problems (Table III). A source incorporation problem was described as "extended," if it incorporated two or more sentences from the original source. The second coding team first determined the type of source incorporation problem independently, then compared the codes and discussed and resolved any disagreements.

This study was determined to be exempt from IRB approval by the UC San Diego Human Research Protections Program, project #120621XX.

Results

In the 255 laboratory reports included in this study, Turnitin® identified, on average, 29 similarity hits to other texts in Laboratory report 1 and 28 matches in Laboratory Report 3. However, after discarding the matches, we deemed insignificant and accounting for Turnitin® misattribution problem (matches to the laboratory manual were misattributed to laboratory reports by other students, see Methods and Table II), we detected an average of two significant source incorporation problems per laboratory report (1.6 in Laboratory report 1 and 2.2 in Laboratory report 3).

Over half of the laboratory reports (53%) exhibited at least one source incorporation problem, and 8% of the laboratory reports contained five or more source incorporation problems (Table IV). Among the 341 source incorporation problems we detected, 41.5% extended beyond a single sentence. The source incorporation problems occurred most often within sections rich in technical details: Materials and Methods and the parts of Results where students wrote about procedures (50.5% of source incorporation problems), or in the Introduction section, where students wrote about the background to the experiments (47% source incorporation problems).

TABLE II *An example of Turnitin® mistakenly attributing the source text to other students' papers, instead of the laboratory manual*

Student's text	*Laboratory manual*
Excerpt from the Introduction: Fire and Mello discovered (**1) that the introduction of double-stranded RNA (dsRNA)** including **coding sequences of a specific gene can disrupt the function of that gene by inducing the** degradation **of** a target **mRNA.**	Basically, they found **that the introduction of double-stranded RNA (dsRNA)** that includes **coding sequences of a specific gene can** specifically **disrupt the function of that gene by inducing the** destruction **of** the **mRNA.**
Excerpt from the Materials and Methods: The *C. elegan* strain used, (**1) strain NL2099-rrf-3, is particularly sensitive to RNAi.** Each plate contained **three to four** large **worms** on them. The (**9) worms were transferred to plates coated with a specific strain of bacteria HT115 (DE3) that** has the **gene for T7 polymerase. T7 polymerase expression is under the control of a *lac* promoter and operator. The HT115 (DE3) bacteria also contains a plasmid** (L440 double T7 vector) (**7) that has an amp resistance gene, and will either contain no RNAi insert (control), or a 800 bp sequence for the unc-22 gene in the polylinker region. Two T7 promoters flank the** (**64) *unc-22*** gene **sequence and control the transcription of the** insert and the **product is** a **dsRNA.**	You will be given untreated plates of gravid *C.elegans* **(strain NL2099-rrf-3—** this strain **is particularly sensitive to RNAi). Three to four worms** will be **transferred to plates coated with a specific strain of bacteria HT115(DE3**) which contains **the gene for T7 polymerase.** The **expression** of the **T7 polymerase is under the control of a *lac* promoter and operator** (so just like the pET vector system). **The HT115(DE3) bacteria also contain a plasmid that has an amp resistance gene, and will either contain no RNAi insert (control), or a 800 bp sequence for the unc-22 gene in the polylinker region.** Transcription of the ***unc-22* sequence** is controlled by **two T7 promoters** that **flank** the insert on both ends, so **the product is a dsRNA**.

Matching text is indicated in bold. Only the relevant parts of the lab manual (Butler and Noree, 2012) are shown. The numbers in parentheses in the left column (1, 9, 7, and 64) that precede the matched text indicate the different sources to which Turnitin® attributed these matches. Match 1 was attributed by Turnitin® to the laboratory manual, while matches 9, 7, and 64 were attributed by Turnitin® to laboratory reports by other students. However, all of the matched text is also present in the laboratory manual (right column, bold), making patchwriting from the laboratory manual a more parsimonious source of the matches. Here and in other tables, we present students' texts without correcting grammatical or spelling mistakes; only the scientific names of organisms (e.g., C. elegans) and of genes were italicized by us.

TABLE III ***Definitions and illustrations of the four different types of source incorporation problems found in students' laboratory reports***

Code	*Definition*	*Student text*	*Source text*
Copying	Verbatim match to an entire sentence (words and structure) in a source (may include minor word substitutions or omissions, such as replacing "will" for "would" or omitting articles or conjunctions, such as "and, but, when, or")	These digested pieces of DNA were then ligated together using T4 ligase **to create recombinant vectors containing pieces of *Vibrio* DNA with Sal I "sticky" ends ligated into cut pGEM with Sal I "sticky" ends.**	**To create recombinant vectors containing pieces of *Vibrio* DNA with Sal I "sticky" ends ligated into cut pGEM with Sal I "sticky" ends.** Source: Lab manual (Butler and Noree, 2012)
Patchwriting	A match to the source that reproduces the original language, but includes synonym substitutions, word/phrase omissions, and sentence restructuring (Howard, *et al*. 2010)	**The genes coding for** a and b **subunits** of **luciferase are *luxA* and *luxB,*** while ***luxC, D,* and *E* genes code for polypeptides that are required for the conversion of fatty acids into the long-chain aldehyde required for the luminescent reaction**…	**The genes coding for** the bacterial **luciferase** enzyme **subunits** which catalyze the bioluminescence reaction **are *lux*A and *lux*B.** The ***lux*C, D, and E genes code for polypeptides** (transferase, esterase, and reductase) **that are required for the conversion of fatty acids into the long-chain aldehyde required for the luminescent reaction.** Source: Lab manual (Butler and Noree, 2012)
Technical Parroting	A match was coded as technical parroting if it met three criteria: 1) It contained repeated material from the lab manual or lecture slides, with essentially little or no change from the original 2) It was found in the Materials and Methods or the Results section 3) It was rich in technical details such as temperatures, names of reagents, concentrations, or volumes	RNA samples were then diluted to a final concentration of 20 ng/ ul using RNase free water, to which a mixture of **master mix from Biorad containing SYBR Green Dye, AmpliTaq Gold DNA Polymerase, dNTPs, reverse transcriptase enzyme, and buffer components were added.**	**Master Mix from Biorad containing SYBR Green Dye, AmpliTaq Gold DNA Polymerase, dNTPS, reverse transcriptase enzyme and buffer components were added.** Source: lab manual (Butler and Noree, 2012)
Falsification	The student falsifies a citation, suggesting the text is from one source when really it is from another. In the example on the right, the student cited the material as having come from Winston *et al*., (2002), but it actually came from the laboratory manual. Note that this falsification is combined with very close patchwriting (almost copying) from the laboratory manual.	**dsRNA** can **move freely from cell to cell in *C. elegans* through a pore formed by a protein, SID-1** (Winston *et al*., 2002).	Amazingly, **dsRNA moves freely from cell to cell in *C. elegans* through a pore formed by a protein, SID-1.** Source: Lab manual (Butler and Noree, 2012)

Matching text is indicated in bold.

TABLE IV *Frequency of the types of source incorporation problems. N = 255 laboratory reports*

Types of source incorporation problems	*Number of source incorporation problems*[a]	*% of total source incorporation problems*[b]	*% extended*[c] *source incorporation problems in this category*[d]	*Number of laboratory reports containing this source incorporation problem*
Patchwriting	201	59%	33%	101 (40%)
Technical parroting	94	28%	78%	64 (25%)
Copying	30	9%	20%	23 (9%)
Falsification	16	5%	19%	13 (5.5%)
Total instances of source incorporation problems	341	100%[e]	41.5%[f]	136 (53%)[g]

[a]*Total number of source incorporation problems in all laboratory reports.*
[b]*Proportion of each type of source incorporation problems, out of the total instances of source incorporation problems (N = 341).*
[c]*Extended problem refers to a source incorporation problem spanning more than one sentence.*
[d]*Proportion of extended source incorporation problems in each category.*
[e]*Note that, because of the rounding, percentages in rows 2–5 add to more than 100%.*
[f]*Proportion of extended problems in the total number of source incorporation problems (N = 341).*
[g]*Note that, because one laboratory report could have more than one type of source incorporation problem, the percentage numbers in rows 2–5 do not add to 53%.*

We categorized source incorporation problems into four types (see Table III for examples):

1. Patchwriting: a match to the source that reproduces the original language but includes synonym substitutions, word or phrase omissions, and sentence restructuring [11]
2. Technical parroting [12]: a match rich in technical details (i.e., volumes and concentrations), reproduced from the laboratory manual or lecture slides with little or no change. Such matches typically occurred in the Materials and Methods and the Results sections of laboratory reports.
3. Copying: a verbatim match to an entire sentence in a source (may include minor word substitutions or omissions)
4. Falsification of a source: a copied or patchwritten text contains a citation that attributes the information to one scientific source (typically, a journal article), while the source of the information is different (typically, the laboratory manual).

Patchwriting and Technical Parroting

The predominant source incorporation problems were patchwriting and technical parroting (Table III), accounting for 87% of such problems (Table IV). Patchwriting, an excessive use of a source's text with the source text modified slightly through word substitutions or deletions [13], contributed 59% of the source incorporation problems (Table IV). Among the instances of patchwriting, some showed a clear (but unsuccessful) attempt to restructure the original sentence and "make it their own," whereas in other instances, this attempt was so minimal, and the use of the source text so extensive, that it bordered on copying (Supplemental Table S2). One-third of the patchwriting problems extended to more than one sentence (Table IV).

A related problem was technical parroting: the repetition of methods, processes, or procedures from the laboratory manual, with little or no change from the original (Table III). The term "technical parroting" is drawn from Moskovitz and Kellogg, who have suggested that when a procedure is already outlined in detail in the laboratory manual, there is "little for [the students] to do but parrot back selected details from the [laboratory] manual" [12]. Technical details, such as volumes of reagents, temperatures, and times of incubation, were abundant in these matches, and the source was almost invariable in the laboratory manual (Table III). Despite the multitude of technical terms, many of which could not be expressed differently, we still considered technical parroting as a source incorporation problem because in these instances, instead of distilling the most pertinent information from the protocol description, students were simply repeating the general instructions of the laboratory manual. Most (78%) of the technical parroting extended to more than one sentence (Table IV).

Since laboratory reports are typically rich in specific terminology and technical details, is it perhaps impossible to avoid substantial similarity to the laboratory manual or to what has been written before by other students? This does not seem to be the case: we found that many students (29%) managed to successfully present the experimental purpose,

definitions, procedures, results, and discussion without yielding any significant Turnitin® similarity matches (Table V).

Copying

Direct copying—exact word-for-word replication of the source material—was found in only 9% of the total instances of source incorporation problems; we found 30 instances of copying in 23 papers written by 20 different students. The sources of copied material included the laboratory manual (33%), another student's laboratory report (30%), journal articles (23%), websites (10%), and a book (3%). The majority of copied text was contained in isolated single sentences (81% of instances of copying) used to supplement background information (21 instances) or methods (9 instances). Examples of copying are shown in Table III and Supporting Information Table S3.

Falsification

Falsification of sources—a citation of a paper that did not contain the information the student attributed to it—occurred in 5% of the instances of source incorporation problems (Table IV) and was found in 13 papers. In all these cases, the student falsely attributed material that came from a website or the laboratory manual to a journal article (see examples in Table III and Supporting Information Table S4). The majority of instances of falsification of sources (13 out of 16) occurred in conjunction with copying or patchwriting. For example, a text might have been patchwritten from the laboratory manual but attributed to a scientific article which in fact did not contain such information (Table III). In the instructions for writing these laboratory reports, students were asked to incorporate information from journal articles. It is likely that falsification of

TABLE V ***Difference between parroting or patchwriting and writing in one's own words***

Source text	*Student text with problematic source incorporation*	*Student text in their own words*
Three to four worms will be transferred to **plates coated with a specific strain of bacteria HT115(DE3)** which **contains the gene for T7 polymerase**. The expression of the T7 polymerase is under the control of a *lac* promoter and operator (so just like the pET vector system). **The** HT115(DE3) **bacteria** also **contain a plasmid that** has **an amp resistance gene, and will either contain no RNAi insert** (control), **or a 800 bp sequence for the *unc-22* gene in the polylinker region**. Source: lab manual (Butler and Noree, 2012)	About **3–4 worms** were put on **plates coated with a strain of bacteria HT115 (DE3)** that **contains the gene for T7 polymerase. The bacteria contained a plasmid that** had **an amp resistance gene and either no RNAi insert or a 800 bp sequence for the *unc-22* gene in the polylinker region.**	On plates of *C. elegans* including bacterial strain HT115(DE3) containing a *T7 polymerase* gene, samples have either plasmids with no RNAi insert (control) or the *unc-22* sequence between the two T7 promoters of a L440 double-T7 vector (experimental). These plates were confirmed to have three to four large worms, with at least two alive.
"***LuxI* and *LuxR* form a 'quorum sensing' regulatory circuit that induces bioluminescence at high cell density**" Source: journal article, Bose *et al.*, 2008, p. 26. "**…system of stimulus and response correlated to population density** " Source: Wikipedia	Luminescent expression in bacteria is dependent on cell density and the ***luxR and luxI* form the quorum sensing regulatory circuit that induces bioluminescence at high cell density.** ***(*****Quorum sensing is a system of stimulus and response correlated to population density*****.)***	The *lux* operon is a very intricate system, that is responsible for the emission of light. …*Lux I* codes for the autoinducer, which when bound by the regulator protein produced by *luxR*, binds to the promoter region to greatly increase transcription of the lux operon. *LuxR* happens to be located directly to the left of the l*ux* operon, and contains its own promoter region.

Left column: source text from which the problematic writing (center column) was derived. Right column: a corresponding part from a laboratory report written by a different student, demonstrating writing in their own words. Matching text is indicated in bold.

sources was an attempt to make it appear as if this requirement was satisfied.

Superficial Use of Primary Literature

A total of 18% of the laboratory reports contained instances of patchwriting or copying that occurred when students attempted to use scientific literature as source material. Frequently, the superficial changes students introduced into the patchwritten text resulted in a loss of the biological meaning of the original text (Table VI). The sentences selected by students for such superficial use were predominantly from a paper's Abstract or the beginning of its Introduction section.

Discussion

Plagiarism or Misuse of Sources?

This study was motivated by the authors' concern about the large numbers of similarity matches detected by Turnitin® in the laboratory reports written by our students. We sought to determine whether these matches were a result of chance alone (i.e., numerous students writing on the same topic and using the same terminology) or whether these matches indicated genuine source incorporation problems: instances of significant similarity between student's laboratory report and another text, such as laboratory manual, website, and journal article. If these matches were indeed source incorporation problems, did they rise to the definition of plagiarism?

Source incorporation problems were present in more than half of the laboratory reports we examined. However, we consider only a small minority of those as plagiarism: copying (14%) or falsification of sources (5%), Most of source incorporation problems we encountered involved patchwriting and "parroting" protocol details from the laboratory manual. While patchwriting and "parroting" can be considered plagiarism [14, 15], we reasoned that students were not fraudulently trying to present the ideas from the laboratory manual as their own, since the laboratory manual's text was available and well known to the instructor. Therefore, we consider patchwriting and technical parroting a misuse of sources, rather than plagiarism.

Our findings about the prevalence of source incorporation problems agree with a multiinstitutional study conducted by Jamieson and Howard (2013) that examined the use of source material in 174 research papers written by freshmen enrolled in writing courses. At least one instance of patchwriting from sources was found in 52% of the papers [16]. Similarly, Flaspohler and colleagues reported that 50% of biology students enrolled in an upper division elective course committed what the authors called "knowing or unknowing plagiarism" when writing evaluative annotations of research articles [17]. The incidence of direct copying we detected was higher than has been previously reported in freshman writing courses: 9% in our study versus 4.3% previously reported [16] perhaps reflecting differences between the contexts of the writing

TABLE VI ***Examples of patchwriting from the primary literature that distorted the meaning of the original text***

Source text	*Problematic student text*	*Notes*
The host organism can use light emitted by bacteria **for attraction of prey, escape from predators or intra species communication** ... From: Czyz *et al.*, 2000 (Introduction section, 2nd paragraph)	Bacterial species use luminescence either for the **attraction of prey, escape from predators or** even **communication** between **species.**	In the original text, it is the host and not the bacteria that use bioluminescence for the described purposes.
A polyclonal antibody raised against an *Escherichia coli* beta-galactosidase-unc-22 fusion protein recognizes **a polypeptide** in nematode extracts **that is between 500,000 and 600,000 daltons and** labels **the muscle A-band** in indirect immunofluorescent microscopy. From: Moerman *et al.*, 1988 (Abstract)	*unc-22* gene yields a polypeptide **that is between 500,000 and 600,000 Daltons** and immunofluoresces at **the muscle-A band.**	In the original article it is the fluorescently labeled antibody that labels unc-22 protein and allows detection. In the student's paper the protein itself immunofluoresces.

Matching text is indicated in bold.

assignments: biology laboratories in this study versus writing courses and Jamieson and Howard's study.

Superficial Use of Primary Literature

In 18% of the laboratory reports, we detected patchwriting from the primary literature, where students cited the source paper, but patchwrote sentences from the scientific articles in ways that very often distorted the meaning of the original sentence. In most instances, only sentences from the Abstract or the beginning of the Introduction sections of the source articles were used, suggesting that students did not read the source articles beyond these sections. These findings are in agreement with a study by Jamieson and Howard (2013) that found that the majority (69%) of citations in 174 papers written in first-year writing courses from a variety of colleges and universities came from the first two pages of the cited sources. Science students also frequently rely on abstracts alone while writing about scientific articles [17].

Howard and colleagues described this phenomenon as "writing from sentences": an attempt to paraphrase or summarize based on just a few sentences from a source [11]. Without the context and deeper understanding of the subject of the article, this task is very challenging for students, leading to a heavy reliance on the original wording of the text and to misrepresentation of the original meaning.

Limitations and Future Directions

The strength of this study is that it examines actual student writing rather than relying on the student self-reports that dominate plagiarism studies (noteworthy exception is Rebecca Moore Howard's Citation Project, found at http://site.citationproject.net/). The ability to generalize our results is limited, since this study was conducted at one institution, in one discipline (biology), and in one type of laboratory report (the expository and traditional laboratory report). Further research in other laboratory science courses is needed to gain a comprehensive view of how undergraduates develop their source incorporation skills.

Our research also has a limitation in common with the research conducted within the Citation Project [16]: while we can identify problems with use of sources, we can only hypothesize as to the causes of such problems. The Council of Writing Program Administrators suggests a number of possible causes of plagiarism and misuse of sources: students may lack the training of appropriate integration of sources, mistakes in integrating sources are expected part of the learning process, college instructors may underestimate the difficulty of such integration [2]. We agree that patchwriting and technical parroting can result from lack of knowledge of writing conventions in biology. It is also possible that students did not choose to spend enough time to carefully read and paraphrase the sources. Interviews with the students whose laboratory reports exhibit source incorporation problems would be very informative in elucidating the underlying reasons of these problems.

Recommendations

Reconsideration of Traditional Laboratory Reports

Alaimo and colleagues [18] and Moskovitz and Kellog [12] argue that the traditional laboratory report is an artificial genre and is not an optimal medium for developing competency in scientific writing. Our study offers evidence to substantiate these claims. In Introduction section of the laboratory report, we asked students to find and write about relevant background literature. This is an important skill since many of our students hope to become physicians or researchers and the ability to read and analyze primary literature has been identified as one of the entry-level competencies for aspiring physicians [19]. However, our findings show that many students instead patchwrote or copied information provided in the laboratory manual. We suggest providing students with less background information in the laboratory manual, so that they explore the relevant background literature.

Technical parroting, another frequent problem, was mostly characterized by a mere repetition of the laboratory manual's information, a "text dump" with limited attempt to understand relevance (no prioritizing, sequencing, or other evidence of deeper understanding). We concur with the argument that laboratory reports can become a more useful and meaningful exercise in the authentic research context: laboratories where students develop protocols and communicate them in a formal written format, so that others can replicate the process [12, 18]. A call for discovery-based research in undergraduate laboratory courses in STEM disciplines was issued by the President's Council of Advisors on Science and Technology [20]. Course-based undergraduate research experiences, or CURE's, offer a promising model for allowing students to engage in authentic research that is also of interest to the scientific community [21, 22]. Indeed, the curriculum of the laboratory course described in our study now includes an authentic research project that has replaced one of the two large projects that engendered the laboratory reports examined here.

This recently implemented course module involves students in hypothesis building and in experimental design. It will be interesting to examine whether such labs will reduce the incidence of mindless patchwriting and parroting from the laboratory manual observed in this study.

In addition, to reduce technical parroting, one of the instructors (E.T.) replaced large-scale, high-stakes laboratory reports with smaller and more frequent assignments in which students practice summarizing procedures, as well as presenting and discussing their results. In such assignments, students are provided with a rubric that includes explicit expectations for summarizing the procedures, as well as with information about grade deduction for parroting the laboratory manual (an example rubric is provided in the Supporting Information, Appendix C). These "mini-laboratory reports" are graded by graduate instructional assistants, who also provide students with feedback on their writing. Our preliminary observations indicate that technical

parroting was reduced in students' writing, following this curriculum change (E.T., unpublished observations).

However, we doubt that authentic research or more frequent, smaller scale writing assignments will solve the problems we see when students attempt to write relying on primary literature. Below, we offer recommendations for dealing with this problem.

Training for Students

Multiple studies have suggested that undergraduates often lack not only the skills of paraphrasing and summarizing but also of conducting an informed literature search and engaging with the ideas in their sources [10, 11, 13, 16, 17, 23–27]. Struggles with source-based scientific writing are to be expected as novices master disciplinary conventions [28], as students grapple with new vocabulary, concepts, scientific writing conventions, and—when writing from research articles—with complex scientific content [16]. Novices often start with patchwriting, advancing toward paraphrasing, and then summarizing as they become more knowledgeable and confident in the discipline [11, 29]. When asked to paraphrase or summarize content from an unfamiliar field, not only students, but also faculty often patchwrite [10, 30]. Therefore, rather than viewing patchwriting as an ethical failure, we suggest it is a problem that needs to be addressed with better-scaffolded instruction in discipline-specific reading and writing.

Such instruction should include training in conducting a literature search, critical reading of scientific articles, and appropriately paraphrasing and summarizing scientific ideas from sources. Training in scientific writing should also include explicit instruction in identifying and avoiding plagiarism [2, 9, 31, 32]. The accompanying manuscript by Yang and colleagues offers a successful approach to addressing these issues in the context of our laboratories. The process of writing about scientific background should include multiple drafts, reviewed by the instructor or appropriately trained teaching assistants and examined via similarity-detection software [2]. Students who struggle with comprehension, paraphrasing, or summarizing should be provided with additional instruction (on reading scientific articles, disciplinary writing, or the ethics of scientific writing) and opportunities to revise their writing. Conversations between instructors and these students could inform a decision regarding which interventions would be most appropriate. Because of its scope, such training should begin early in the undergraduate program.

Changes in Turnitin®

Turnitin® similarity reports were helpful in the initial identification of possible source incorporation problems, although some current features of Turnitin® made the determination of the extent of plagiarism and the source of patchwriting difficult in our context. For example, we found many cases in which similarities that were clearly derived from one text available to all students (the laboratory manual) were identified by Turnitin® as small individual matches to a large number (sometimes dozens) of student papers. This presents a substantial difficulty for instructors in interpreting Turnitin® results: they are left to puzzle if such matches are evidence that the student is plagiarizing or are mere coincidences.

Interpretation of Turnitin® results would be simplified if instructors could indicate which texts (e.g., laboratory manuals and scientific papers) Turnitin® should consider the "preferred," most likely source material. Alternatively, the Turnitin® algorithm should be changed so that it can seek the most likely source of plagiarism. It would also be very useful for instructors to have a means for indicating phrases that Turnitin® should ignore (e.g., terminology and technical definitions) because there are no other ways to express the information. Such options are currently unavailable on Turnitin® (to our knowledge).

Faculty Training in Using Turnitin®

Based on our findings, we recommend that laboratory instructors should be trained to use the existing version of Turnitin® in the following ways:

1. Before students submit their laboratory reports to Turnitin®, the instructor should upload the laboratory manual or instructions provided to students so that it will become "the oldest" and thus the "original" source for the Turnitin® algorithm.
2. After the student reports are submitted, the instructor should review each Turnitin® originality report to assess which identified matches are of concern (e.g., a match to a journal article or extensive matches to another student's paper). We recommend using the "text-only" version of the originality report as it breaks down the suspected sources rather than showing all "student papers" together as one source.
3. If patchwriting or technical parroting problems are detected, we recommend addressing them first as writing problems rather than as plagiarism.
4. Copying and/or falsification of sources should be considered plagiarism and dealt with according to the institution's procedures for reporting academic integrity violations.

We hope that our recommendations will help instructors who use Turnitin® to recognize and differentiate between instances of plagiarism and misuse of sources. Such differentiation will help them in a difficult task of meeting the needs of their science program to conduct honest assessments and the needs of students who at times are facing expected challenges in learning scientific writing.

Acknowledgments

We are grateful to the Biology undergraduate students for their participation in this study. We thank our colleagues Melissa Micou, Lisa McDonnell, and Cindy Gustafson-Brown for their critical reading of this manuscript and for their insightful comments.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

[1] Carter, M. (2007) Ways of knowing, doing, and writing in the disciplines. Coll. Compos. Commun. 58, 385–418.

[2] The Council of Writing Program Administrators. Defining and Avoiding Plagiarism: The WPA Statement on Best Practices [cited 2018 Nov 18]. Available at: http://wpacouncil.org/positions/WPAplagiarism.pdf.

[3] Walker, J. (1998) Student plagiarism in universities: what are we doing about it? High. Educ. Res. Dev. 17, 89–106.

[4] Macdonald, R., Carroll, J. (2006) Plagiarism—a complex issue requiring a holistic institutional approach. Assess. Eval. High. Educ. 31, 233–245.

[5] Colvin, B. B. (2007) Another look at plagiarism in the digital age: is it time to turn in my badge? ERIC. 35, 149–158.

[6] P. Wasley Antiplagiarism software takes on the honor code - the chronicle of higher education. (2008) Chron. High. Educ. 54 n25 pA12

[7] Zwagerman, S. (2008) The Scarlet P: Plagiarism, Panopticism, and the Rhetoric of Academic Integrity. Coll. Compos. Commun. 59, 676–710. Available at: http://www.jstor.org/stable/20457030.

[8] Del Carlo, D., Bodner, G. (2006) Dishonesty in the biochemistry classroom laboratory: a synthesis of causes and prevention. Biochem. Mol. Biol. Educ. 34, 338–342.

[9] Holt, E. A. (2012) Education improves plagiarism detection by biology undergraduates. Bioscience 62, 585–592.

[10] Roig, M. (1999) When college students' attempts at paraphrasing become instances of potential plagiarism. Psychol. Rep. 84, 973–982.

[11] Howard, R. M., Serviss, T., Rodrigue, T. K. (2010) Writing from sources, writing from sentences. Writ. Pedagog. 2, 177–192.

[12] Moskovitz, C., Kellogg, D. (2011) Inquiry-based writing in the laboratory course. Science 332, 919–920.

[13] Howard, R. M. (1992) A Plagiarism Pentimento. J Teach Writ 11, 233–246.

[14] Hacker, D., Sommers, N. (2015) A Writer's reference with writing in the disciplines. Eighth Macmillan Higher Education, Bedford/St. Martins, Boston.

[15] UW-Madison. Writing center writer's handbook. 2018. [cited 2018 Nov 18]. Available at: https://writing.wisc.edu/Handbook/.

[16] Jamieson, S., Howard, R. M. (2013) Sentence-mining: uncovering the amount of reading and reading comprehension in college writers' researched writing. In: McClure R., Purdy J. P., Eds. The New Digital Scholar: Exploring and Enriching the Research and Writing Practices of NextGen Students, American Society for Information Science and Technology, Medford, NJ. pp. 111–133.

[17] Flaspohler, M. R., Rux, E. M., Flaspohler, J. A. (2007) The annotated bibliography and citation behavior: enhancing student scholarship in an undergraduate biology course. CBE—Life Sci. Educ. 6, 350–360.

[18] Alaimo, P. J., Bean, J. C., Langenhan, J. M., Nichols, L. (2009) Eliminating lab reports: a rhetorical approach for teaching the scientific paper in sophomore organic chemistry. The WAC J 20, 17–32.

[19] American Association of Medical Colleges and Howard Hughes Medical Institute (AAMC-HHMI). (2009) Scientific foundations for future physicians. Washington, DC, and Chevy Chase, MD. Available at: https://www.aamc.org/download/271072/data/scientificfoundationsforfuturephysicians.pdf. Accessed on March 21, 2019.

[20] President's Council of Advisors on Science and Technology (2012) Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering and Mathematics, U. S. Government Office of Science and Technology, Washington, DC.

[21] Corwin, L. A., Graham, M. J., Dolan, E. L. (2015) Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. CBE—Life Sci. Educ. 14, es1.

[22] E and M (2015) National Academies of Sciences. Integrating Discovery-Based Research into the Undergraduate Curriculum, National Academies Press, Washington, DC.

[23] Hull, G. M. (1989) Rethinking remediation: toward a social-cognitive understanding of problematic Reading and writing. Writ. Commun. 6, 139–154.

[24] Wells, D. (1993) An account of the complex causes of unintentional plagiarism in college writing. WPA: Writ Program Adm 16, 59–71.

[25] Horning, A. S. Across the disciplines reading across the curriculum as the key to student success. (2007). [cited 2016 Nov 14]. http://wac.colostate.edu/atd/articles/horning2007.cfm.

[26] Horning, S. (2010) A potential solution to the plagiarism problem: improving reading. J Teach Writ 25, 143–175.

[27] Szymanski, E. A. (2014) Instructor feedback in upper-division biology courses: moving from spelling and syntax to scientific discourse. Across Discipl. 11. Available at: http://wac.colostate.edu/atd/articles/szymanski2014.cfm. Accessed on January 14, 2016.

[28] Colomb, G. G. (1988) Disciplinary "Secrets" and the Apprentice Writer: The Lessons for Critical Thinking. Institute for Critical Thinking Resource Publication Series 1 No. 6.

[29] McGowan, U. (2005) Academic integrity: An awareness and development issue for students and staff. J. Univ. Teach. Learn. Pract. 2, 48–57.

[30] Roig, M. (2001) Plagiarism and paraphrasing criteria of college and university professors. Ethics Behav. 11, 307–323.

[31] Franklyn-Stokes, A., Newstead, S. E. (1995) Undergraduate cheating: who does what and why? Stud. High. Educ. 20, 159–172.

[32] Gehring, K. M., Eastman, D. A. (2008) Information fluency for undergraduate biology majors: applications of inquiry-based learning in a developmental biology course. CBE—Life Sci. Educ. 7, 54–63.