CHAPTER 5

Sentence-Mining: Uncovering the Amount of Reading and Reading Comprehension in College Writers' Researched Writing

Sandra Jamieson and Rebecca Moore Howard[1]

The *Writer's Guide and Index to English*, a college writers' handbook in wide circulation at the middle of the last century, articulates an ideal for students' work from sources that endures today:

> A student—or anyone else—is not *composing* when he is merely copying. He should read and digest the material, get it into his own words (except for brief, important quotations that are shown to be quotations). He should be able to *talk* about the subject before he *writes* about it. Then he should refer to any sources he has used. This is not only courtesy but a sign of good workmanship, part of the morality of writing. (Perrin 1959, 636) [**Au: Emphasis in original?**]

This brief statement buried deep in an antiquated writers' handbook is remarkable for several reasons, not least of which is its crisp, accessible presentation of a complex truism of academic writing. The idea that writers must be able to "*talk* about the subject" is at the heart of the notion of writing as "conversation" that is repeated in scholarly articles, outcomes statements, and the language of current pedagogy. While prewriting activities use writing as a means of discovery, that process of discovery is embraced by many as a way to enable students to be able to "talk about" their topic before they begin to construct arguments and papers. Few of us would feel the need to *say* this today, but studies of students' researched papers suggest that we should.

Perrin's (1959, 636) statement is remarkable because of its association of "get[ting] it into his own words" with understanding—"digesting"—the source. The passage excludes copying from the realm of composing. When one copies, says the *Writer's Guide*, one is not composing. One is merely copying. Note that when he speaks of "copying," Perrin is not talking about unattributed copying, but *all* copying, including attributed quotation. When one copies, he says, one is not talking about the subject, but merely transcribing others' talk. This claim is complicated. Some academic disciplines value the transcription of others' talk, calling for quotation of significant text rather than paraphrase. Others reject quotation, calling for a synthesis of ideas and findings rather than an emphasis on specific words. Yet across this difference is a shared desire for students to *understand* their sources. If students are quoting or paraphrasing one or two sentences at a time, they are not "digesting" the ideas in the source and using those ideas to compose papers and reports of their own. They are, in Perrin's terminology, copying.

The field of college writing instruction values and teaches the skills of paraphrase and summary—the "digesting" of texts considered by Perrin to be integral to composing from sources. Faculty outside of writing studies also value these writing skills in discipline-specific and general student writing. Conducting cross-disciplinary research on the ways college instructors experience intellectual property and represent

it to their students, Lise Buranen and Denise Stephenson describe a chemistry instructor who encourages his students to paraphrase rather than quote, in part to increase their understanding of the source text (2008, 73). The belief that the act of paraphrasing or summarizing helps writers understand their sources is articulated in faculty development work and guides to research, and it is frequently asserted in writing studies scholarship and textbooks. It seems to be a disciplinary or even academic given; nowhere have we seen a compositionist challenge this tenet. We have ourselves promoted the value of summary and paraphrase in our teaching, our work as writing program administrators, and, beginning as early as 1992, our scholarship (Howard 1992).

Our experiences as teachers and administrators of college writing lead us to fear that Perrin's (1959) last principle—that copying is not composing—is being obscured by our current culture of plagiarism hysteria. In their rush to discourage plagiarism, college instructors across the disciplines may be so concerned about students' successful enactment of the mechanical process of *acknowledging* copying that the rhetorical and intellectual dimensions of cross-textual work fade into the background. And when those instructors assess student writing, the result may be that students are rewarded for successful citation out of proportion to the rhetorical and intellectual quality of their texts. Instructors may not always be noticing whether or how much students are, in Perrin's formulation, copying from sources instead of composing from them.

In order to change this dynamic, we first need to know how much students actually use paraphrase and summary in their writing from sources. We also need to know how much they patchwrite, which the Citation Project and others define as working too closely with the language and syntax of the source when they attempt to paraphrase.[2] If we are to explore student *understanding* of texts, we need to see what they do with their sources. Working from multi-institutional research known as the Citation Project, this chapter provides data that begin to answer that question.

Background

A study of student source use by Rebecca Moore Howard, Tricia Serviss, and Tanya K. Rodrigue (2010) found that students worked with sources at the sentence level instead of representing the larger ideas in the source through summary. Expanding on Diane Pecorari's study (2003) of the ways nonnative speakers of English incorporate sources, they explored the extent to which college students' researched writing incorporated four source-use techniques: copying, patchwriting, paraphrasing, and summarizing. Their study found no summary in the 18 researched papers analyzed. It also found that within those papers, it "is consistently the sentences, not the sources, that are being written from" (Howard, Serviss, and Rodrigue 2010, 189). This research, based at one institution, prompted us to ask more questions and design a multi-institutional quantitative study of student papers produced in the first-year writing course or course sequence at 16 U.S. colleges and universities. Those institutions were chosen to represent the entire geography of the country and its most common types of institutions.

As with the single-institution study, the multi-institutional analysis found that the most common form of citation was direct quotation (46 percent of all of the citations in the 174 papers in this study), followed by paraphrase (32 percent) and patchwriting (16 percent). Only 6 percent were summary—even if we define that term generously. In other words, 94 percent of the citations were created by students working with their sources at the sentence level and not demonstrating that they had "digested" what they read. But these data were not, in fact, our most compelling findings. In addition to not summarizing their sources, our data suggest that many of the students whose papers we analyzed may not even have read beyond the first few pages of the source.

Our research is based on some essential principles. The first is that as scholars and administrators we need to base our claims about what students do on solid data. The contemporary obsession with plagiarism

is possible because those who report and repeat it are working from experience, anecdote, and over-generalized claims about student integrity. For example, it seems logical to assume that the expansion of the internet would increase student plagiarism, especially if one is predisposed to believe that students will cheat if given the opportunity. Yet we do not have data about the extent of plagiarism before the internet, so we have nothing to compare with post-internet plagiarism. All we know is that the internet makes it easier to catch plagiarists. Without meaningful data, anecdote and beliefs about students will continue to dominate the conversation. Similarly, although writing teachers spend considerable time teaching summary and paraphrase, and alone or with librarians emphasize information literacy and source retrieval, we could not evaluate our success until we had local and multi-institutional data to tell us how our students used that information.

The second principle of the Citation Project is that to be meaningful, data needs to come from a wide variety of institutions. Those institutions need to be different in kind and geographical location. While data from single institutions are invaluable for assessment and as pilot research to allow the formulation of more nuanced questions and more efficient data processing, they cannot be used to make broad generalizations about what students do or do not do. In order to be able to speak meaningfully about the trends in student writing in the United States, we undertook to compile a data-based portrait of how students in writing courses work with their sources. That portrait is drawn from the work of 174 students at 16 colleges and universities from a wide geographical distribution in the U.S. Participating institutions are located in 12 states (Alabama, Colorado, Georgia, Idaho, Indiana, Kansas, Massachusetts, New Hampshire, New Jersey, New York, Texas, Washington) and include community colleges, Ivy League institutions, liberal arts colleges, religious colleges, private colleges and universities, and state colleges and universities. The goal of the Citation Project is to collect and share multi-institutional data that will inform the work of scholars, teachers, and administrators and the design and assessment of pedagogies and policies.

The Citation Project also works on the principle that researchers in the field of writing studies must adopt or adapt methods of quantitative analysis already established in other fields if they seek to develop an overall understanding of what students do when they write.[3] Since Chris Anson's call for data-based research in writing in his keynote address at the Council of Writing Program Administrators conference in 2006, the field has seen an increase in this kind of research, and we were also motivated by that speech (published in expanded form in 2008). It is still somewhat unusual to attend sessions at conferences where scholars are presenting data generated by SPSS (Statistical Package for the Social Sciences; the leading computer program for social science-based statistical analysis), but this trend is increasing and we are no exception. Our research uses citation context analysis, a set of research methods established in the fields of applied linguistics and information studies, and adapts it to the field of writing studies.[4] We also employ qualitative and rhetorical methods with which our field is more familiar. Using qualitative data to present an overall picture and generate questions and using quantitative data to explore those questions[5] allows deep and nuanced understanding. And as the qualitative analysis generates more questions, the cycle repeats.

Methods

Source and Paper Coding

Phase I of our research focused on the researched writing produced in standard first-year writing courses. We invited participating institutions to send us at least 50 researched papers of seven or more pages written in at least four sections of first-year writing taught by at least three different instructors. Those papers were randomized; then we rejected any that were too short or whose sources we could not find. We gathered papers from three institutions in Spring 2008 and the remaining 13 in Fall 2009 and Spring 2010, reporting our findings from those first three institutions in a number of presentations while we collected and analyzed the remaining papers. This was a very labor-

intensive process that included a team of 25 compositionists, both faculty and graduate students, working alone and in pairs.[6]

Our database includes 50 pages of student writing—between 1,000 and 1,150 lines of prose—from each institution. So between them, the 16 participating institutions gave us 800 pages of student research, a total of 17,600 lines of prose. In most cases, those 50 pages came from pages two through six of each of 10 papers. By beginning on the second page, we were able to focus on the source use in the body of the paper where the students were most frequently engaging with researched material. The coded pages in each set of papers from each campus included an average of 119 citations to 58 sources, which combined to give us an overall total of 1,911 citations to 930 sources. We found those sources,[7] coded them by type, and then coded the ways they were used in the student papers. In the interest of space, the specific methods we use to code papers and sources are described only briefly here; however, they are available in much more detail on our website (www.citationproject.net), where our training materials and handouts may also be found.[8] Because the citations we studied came from only 10 to 12 papers per institution, our findings for each institution are of limited use when taken alone; however, our project was to look for patterns across institutions. If we found those patterns and if the data from each institution fit the general pattern, the data would be useful locally and also as a way to trace overall trends.

Our data concerning sources selected and used will be published elsewhere as part of our analysis of the information literacy practices of the students in our study. (All publications are listed at www.citationproject.net.) This chapter focuses on the ways students incorporated information from their sources into their papers. The descriptions we used for each of these types of source uses were described for paper-coders in Table 5.1.

While it is easy to define what we mean by "copied" and "quotation," the other three terms are not so straightforward. In 1993, Howard defined patchwriting as "[c]opying from a source text and then deleting some words, altering grammatical structures, or plugging

Table 5.1 Types of Source Use, From "Instructions for Paper Coders"

Passage copied exactly, but not marked as quotation	Exact copying with a citation but no quotation marks or indentation to signal that this is quoted material. (May include minor errors in transcription.)
Passage copied exactly, marked as quotation	Exact copying with a citation and quotation marks or indentation to signal that this is quoted material. (May include minor errors in transcription and errors in citation as long as the copied material is identified as such.)
Passage patchwritten	Restating a phrase, clause, or one or more sentences while staying close to the language or syntax of the source.
Passage paraphrased	Restating a phrase, clause, or one or two sentences while using no more than 20 percent of the language of the source.* Paraphrase does not necessarily involve significant reduction in length.
Passage summarized	Restating and compressing the main points of an entire text or at least three or more consecutive sentences in the text, reducing the summarized passage by at least 50 percent and using 20 percent or less of the language from that passage.*

*NOTE: This 20 percent does not include accurate synonyms, articles, prepositions, proper names, technical terms, or other keywords. This 20 percent does include words whose morphology is changed (a change in verb tense, for example).

in one-for-one synonym-substitutes" (233); however, this definition implies an intentionality that we have not always found to be the case. For this research, we set out to define the term as neutrally as possible. We felt compelled, however reluctantly, to quantify paraphrase and summary. We did not find ourselves counting words very frequently, though. Passages that were patchwritten generally used significantly more than 20 percent of the source material (more than 50 percent most of the time).

In contrast, because our definition of summary requires a reduction by 50 percent of the material in at least three consecutive sentences, passages of summary generally include significantly less than 20 percent of the language of the source. Brown and Day (1983) report on six "rules" that writers follow when summarizing: Two involve deletion of material from the source text; two involve generalizing from specifics in the source text; and two require invention of sentences that capture the gist of one or more paragraphs (178). Although they were not part of our coding guidelines, these rules did seem to be at play in text coded as summary.

In most cases, patchwriting can be identified with as much ease as can summary once one has read the original source. An example from a student paper in the study demonstrates this in Table 5.2, with marginal coding indicating how the source is being used. In each text, words copied directly from the source are underlined with a single line and word substitutions are indicated with wavy underline.

The student paper from which these extracts were taken includes three citations to material from five paragraphs of a web page produced by NORML, an organization that describes itself as "working to reform marijuana laws" (www.norml.org). The section of the NORML website accessed by the student includes a link to a downloadable PDF of a 57-page report, which is summarized on the pages the student cites; however, the citations clearly reference this website rather than the article. The student works sentence-by-sentence through each of the paragraphs on what prints out as the second page of the three-page source. Two of the three citations to this source are included in Table

Table 5.2 Sample From Source Text and Student Paper

	Student text (page 6 of paper)	**Source text (page 2 of source)**
(1) Paraphrase	Evidence of a jump in interest can be seen in a jump from **258 journal articles** that were **published** in **1996** on the subject of **cannabis**, to **over 2,100 studies** that were published in **scientific** journals in **2008** (Recent Research on Medicinal Marijuana).	A keyword search using the terms "cannabis, 1996" (the year California voters became the first of 14 states to allow for the drug's medical use under state law) reveals just **258** scientific **journal articles published** on the subject during that year. Perform this same search for the year **2008,** and one will find **over 2,100 published** scientific **studies**.
(2) Patchwriting	Most importantly, **investigators are** now **studying the anti-cancer properties of** cannabinoids. There is an increasing amount **of preclinical and clinical data** that conclude **that cannabinoids** stop **the spread**ing **of specific cancer cells** through **programmed cell death** and the prevention of the forming of **new blood vessels** (Recent Research on Medicinal Marijuana).	**Investigators are** also **studying the anti-cancer activities of** cannabis, as a growing body **of preclinical and clinical data** concludes **that cannabinoids** can reduce **the spread of specific cancer cells** via apoptosis (**programmed cell death**) and by the inhibition of angiogenesis (**the** formation **of new blood vessels**).
STUDENT CITATION: "Recent Research on Medical Marijuana." NORML. April 1, 2009. www.norml.org/index.cfm?Group_ID=7002.		

5.2. The third is another example of patchwriting on the same page of the student paper.

The material in the first block of student text in Table 5.2 meets our definition of paraphrase ("Restating a phrase, clause, or one or two sentences while using *no more than 20 percent of the language of the source*"). Although this sentence follows the order of the two sentences in the source text and includes some of the same words, the information is reproduced in one sentence that uses original language. The words that are reproduced are mostly single words and many are specific terms, such as "journal article" and "scientific."

The second extract in Table 5.2 is taken from the next paragraph of the student paper. If we compare the first extract with the second, which we code as patchwriting, we can see the difference between these two ways of incorporating source material. In this second passage of student text, 26 of the 41 words in the source sentence have been reproduced exactly, and another seven have been replaced by synonyms or closely related terms ("cannabis" is replaced by "cannabinoids," and "growing body" with "increasing amount," for example). While some words and phrases have been omitted, the student text follows the same order as the source text and does not add anything original to the sentence or the presentation of the information. This fits our definition of patchwriting: "Restating a phrase, clause, or one or more sentences while staying close to the language or syntax of the source." In addition to repeating words and phrases, the student sentence follows the overall shape of the passage from the source.

Even if the sample of patchwriting in Table 5.2 had been rewritten into a successful paraphrase, it would still be working from just one sentence of the source. We would not, though, be able to see that if we did not read the source material and then track how the student used it.

Inter-Coder Reliability

Coders were placed randomly into pairs so no two coders worked together on all of the papers from a single institution (and at least one of the two coders was from an institution other than the one whose

papers were being coded). Data from their coding was entered into a spreadsheet for each paper, and then coders convened to review their coding and recode as needed, until consensus was reached. Then the information was added to the source-coding information in the SPSS database (PASW Statistics 17).[9]

Where it occurred, variation tended to come from a form of halo effect: Coders sometimes "gave the benefit of the doubt" to otherwise well-written papers and coded passages as paraphrase rather than patchwriting, or summary rather than paraphrase.[10] We found ourselves wanting the students to do well—a very different experience than we have when we set out to "catch plagiarism." Once we became aware of this tendency, we adjusted for it and the process of calibration corrected any potential miscoding by requiring coders to "report the evidence, not a rating" as recommended by those who have studied the effect (Thorndike 1920, 29). The lead researchers blind-coded sources and papers to further ensure inter- and intra-coder reliability and very rarely disagreed with a classification in the final, calibrated data.

Findings

The Papers

The majority of the papers in our database are first-year writing research papers with an argumentative thesis in the introduction and sources used to construct and support that thesis. In their study of handouts for research assignments collected from 28 colleges and universities, Alison Head and Michael Eisenberg (2010) found that "although the topics vary, the assignments consistently demand inquiry, argument, and evidence" (2) with 83 percent requiring students to "write a paper that provides supportive evidence from outside sources" (7).

We did not ask institutions to provide the assignments to which the papers we coded responded, but based on our analysis of the papers, we hypothesize that if we had done so, our findings would be similar to Head and Eisenberg's. Only 54 percent of the assignments in Head

and Eisenberg's sample left the students to select their own topic, but their sample came from faculty and courses from across the curriculum (6). Given the range of topics in the papers submitted from each of the 16 institutions, we believe that the majority of students in our sample selected their own topics.

The Data

Our first research question was focused on Perrin's (1959) claim that a writer "should read and digest the material, [and] get it into his own words (except for brief, important quotations that are shown to be quotations)" (636). How frequently is it the case that students "get it into [their] own words"? How many times do they choose to paraphrase or summarize their sources as they develop a researched paper, and how often does the paraphrase fall short and become patchwriting instead? Our research did not ask whether students made wise decisions, or why they made the choices they did. We simply coded and counted incidences of each. The data in Table 5.3 show the frequency of each kind of citation among the 1,911 citations we coded.

Reading the table row by row, one quickly sees that when these 174 students cited exact copying, they usually marked it as quotation, either with block indenting or with quotation marks. Only 4 percent of the 1,911 citations were to direct copying not marked as quotation, whereas 42 percent of the citations were to direct copying marked as quotation. Regardless of whether the omission of quotation marks was accidental, what we see is that 46 percent of the students simply tran-

Table 5.3 Analysis of Source Use in 1,911 Student Citations[1]

Predominant use of source material within the citation		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Copy without quotation marks	83	4.34	4.34	4.34
	Copy with quotation marks	793	41.50	41.50	45.84
	Patchwriting	306	16.01	16.01	61.85
	Paraphrasing	609	31.87	31.87	93.72
	Summary	120	6.28	6.28	**100.0**
	Total	**1,911**	**100.0**	**100.0**	

scribed the words of others. A further 32 percent of all of the citations were paraphrased, and 16 percent were patchwritten. Adding these to the percentage of citations that were to quoted material, we see that 94 percent of the 1,911 citations were written from isolated sentences in the source texts. Only 6 percent of the citations were to three or more sentences that the student writer had summarized.

The data in Table 5.3 present overall patterns of source use within the 1,911 citations; however, these numbers do not tell us how many individual papers included each type of source use—which was our second research question. We answered this question by analyzing individual papers, and that analysis reveals a slightly different pattern. The data in Table 5.4 show how many of the 174 papers included at least one example of each type of source use in the sample coded.

We only coded five pages in each paper, so there may have been other types of source use in parts of each paper that we did not code. This means we cannot say categorically that something did *not* occur *in the paper*—only that it did or did not occur in the sample we coded. With that caveat, we see a distinct contrast between the frequency of each type of source use in the 1,911 citations and the frequency within each paper.

Table 5.4 Analysis of Source Use in Each of the 174 Student Papers

Type of Source Use Occurring at Least Once in the Paper		Frequency			Percent		
		Occurs at least once in pages 2–6	Does not occur in pages 2–6	**Frequency Total**	Occurs at least once in pages 2–6	Does not occur in pages 2–6	**Percentage Total**
Valid	Copying not marked as quotation	33	141	**174**	19.0	81.0	**100**
	Copying marked as quotation	159	15	**174**	91.4	8.6	**100**
	Patchwriting	91	83	**174**	52.3	47.7	**100**
	Paraphrasing	135	39	**174**	77.6	22.4	**100**
	Summarizing	71	103	**174**	40.8	59.2	**100**

Table 5.3 reveals a total of 120 incidences of summary in the 1,911 citations; however, Table 5.4 shows that only 71 of the papers (41 percent) included *any* incidences of summary, and of the 103 that included no summary, 18 included no paraphrase either, although seven of them included patchwriting—failed paraphrase. The remaining 11 papers depended exclusively on copying in the pages we coded. Although only 11 papers contained no source use other than quotation, the vast majority, 159 of the 174 papers (91 percent), included at least one quotation. The majority of papers also included at least one incidence of paraphrase (78 percent), but a little over half (52 percent) included patchwriting. Of the students who patchwrote, the majority also paraphrased at least once.

If 41 percent of the papers include at least one summary and 78 percent include at least one paraphrase, we might conclude that the students in our sample are engaging with the material, after all. However, other data complicate this interpretation. Our third question asked where in the source students found the material they cited (see Table 5.5).

Table 5.5 Page in Source From Which the Cited Material Is Drawn

Page in source from which material is cited		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Page 1 of the source	885	46.31	46.31	46.31
	Page 2 of the source	443	23.18	23.18	69.49
	Page 3 of the source	151	7.90	7.90	77.39
	Page 4 of the source	100	5.23	5.23	82.62
	Page 5 of the source	73	3.82	3.82	86.44
	Page 6 of the source	48	2.52	2.52	88.96
	Page 7 of the source	31	1.62	1.62	90.58
	Page 8 of the source and beyond	180	9.42	9.42	**100.00**
	Total	**1,911**	**100**	**100**	

The majority, 46 percent of the students' 1,911 citations, come from page 1 of the source. Adding in page 2 takes this percentage up to 69 percent, and a full 83 percent of all of the citations came from one of the first four pages of the source cited—regardless of the length of the source. Only 9 percent of the citations refer to material from page 8 or beyond in the source. Taking this finding into account casts doubt on how engaged the student writers were with the sources they were citing.

Discussion

Misused Source Material—Incorrectly Quoted or Patchwritten Passages

Of the 1,911 citations we studied (Table 5.3), only 4 percent were to material that was cited and copied but not marked as quotation; however, when we look at the 174 papers themselves (Table 5.4) we see that this phenomenon is quite widespread. A total of 19 percent of all of the papers include at least one incidence of direct copying that was cited but not marked as quotation. Similarly, Table 5.3 reveals that within the 1,911 citations, 16 percent were patchwritten from the source; however, as we see in Table 5.4, a total of 52 percent of the 174 papers included at least one incidence of cited patchwriting within the pages we coded. In all, over half of the papers (56 percent), a total of 98 of the 174 papers, included at least one instance of either incorrectly marked quotation or patchwritten prose, and 26 (15 percent) of them included both. These two ways of incorporating source information are designated at best as misuse of sources, and at many institutions they are classified as plagiarism.[12]

This phase of the Citation Project research works only with decontextualized textual artifacts, so we cannot yet report on student intentions. Our hypothesis, though, is that when writers cite patchwritten material, they are attempting to produce paraphrase. Similarly, we suspect that most student writers who cite a source but omit quotation marks are not intending to deceive. Regardless of intentions, the fact

that over half of the students reproduced the ideas of the source in a copied or patchwritten passage that they cited but did not mark as quotation should give us pause. It suggests that policies defining these forms of source use as plagiarism may need to be revised or at least revisited; the textual evidence suggests that the students were not writing well from their sources, but not that they were attempting to claim authorship of passages they did not themselves compose. The difference between unsuccessful writing from sources and academic dishonesty is an important one.

Data-Mined Source Material—Quoted and Paraphrased Passages

When we focus on academic integrity as the gold standard for assessing students' use of sources, we spend less time asking what is happening in student papers that use sources correctly. The cumulative percent column of Table 5.3 raises a different issue, one that we consider more significant than *mis*use of sources. Within the 1,911 citations, 46 percent are to passages that incorporate source material by simply transcribing those sources. In Perrin's (1959) terms, nearly half the time the students were not *composing* from sources.

Quotation holds an essential place in academic discourse, bringing multiple voices to bear on the topic at hand, respecting the precise articulation of a source. We use quotation extensively in this chapter. Quotation does not, however, reveal how much the citer has engaged with the cited text. When a writer only copies from sources, the reader does not necessarily know whether or how well the source has been read. And this is a key question in assessing students' writing from sources.

The use of paraphrase in pedagogy dates back at least to Erasmus (Corbett 1971), and although 78 percent of the 174 students paraphrased at least once in the part of the paper we coded (Table 5.4), paraphrase occurred far less frequently than copying, with only 32 percent of the 1,911 citations being successful paraphrases (Table 5.3). Even if we combine the percentage of successful paraphrase (32 percent) with

unsuccessful paraphrase—patchwriting—(16 percent), we are still left with less than half of the citations reflecting the kind of intellectual intensity David Maas (2002) describes as central to paraphrase. Further, if we review the numbers in the cumulative column of Table 5.3 again, we see that in 94 percent of these 1,911 citations the students were sentence-mining. Copying, paraphrasing, and patchwriting all work from isolated sentences. Only summary works beyond the sentence level.

Digested Source Material—Summary and Paraphrase

In their textbook *Writing Analytically*, David Rosenwasser and Jill Stephen (2006) go so far as to assert, "Summary is the standard way that reading—not just facts and figures but also other people's theories and observations—enters your writing" (117). Judging from the Citation Project findings, Rosenwasser and Stephen are, like Perrin (1959), articulating an ideal rather than describing students' practice. Summary accounts for only 120 (6 percent) of the 1,911 citations (Table 5.3). While it is true that 71 of the 174 students (41 percent) summarized at least once in their papers (Table 5.4), most of them did so *only* once. Using Perrin's terminology, only 41 percent of the papers showed evidence that the student had "digested" *any* of the ideas of the source by summarizing them. It is important to remember that "summary" here can mean something as small as "summary of three consecutive sentences." It also includes one-sentence general plot summaries of works of literature that may have been read for the class. Even with that expansive definition of "summary," we found only 120 incidences of it in 800 pages of student-researched writing (Table 5.3).

Location of Cited Material Within the Source

When we saw the data in Tables 5.3 and 5.4, we wanted to think that surely they did not reflect the best of the students' abilities. Surely, far more often than these data show, the students *did* understand the source and simply weren't demonstrating it by paraphrasing or summarizing. One can engage with the entire source even if one only

quotes from it; however, in many such cases we would expect those quotations to be taken from strategic places from within the text. Table 5.5 challenges that optimism. Not only are students deciding to use quotation to incorporate the majority of their source material, but those quotations usually come from the first or second page of the source. Of the 1,911 citations, 46 percent are to the first page of the source, and a further 23 percent to the second page (Table 5.5).

As with our other data, this finding does not prove that students are not reading the entire source. The first two pages of most academic texts provide some form of summary of the material to follow in the form of an abstract or set of introductory paragraphs that include a thesis or findings to be discussed. In this chapter, we have quoted or paraphrased material from the first page of some of our sources, a notable example being our footnote describing the halo effect in research. In most cases, though, we also reproduce material from elsewhere in the source. To provide only a series of thesis statements or major findings is to fail to provide nuance; readers do not know how the thesis was reached, what constraints surround it, or what role it played in the argument of the source. When students do not include that information, at the very least they reveal that they do not understand its significance. We suspect that this lack of understanding may be at the heart of the problem. While some students may not understand what they read, others may simply not understand what will be gained from reading an entire source, when all the "evidence" they need is right there in the introduction. In other words, our data may be revealing that students do not know how to read academic sources or how to work with them to create an insightful paper.

Our data reveal this tendency to sentence-mine from the first two or three pages from each source text regardless of the overall length of that source. While two of the 174 papers do provide quite extensive summaries of an article that is more than six pages in length (one in each paper), and a few more provide plot summaries of works of fiction, very few of the papers quote or paraphrase from several different pages in one source or draw on one or more sources throughout.

Conclusion

When 94 percent of the citations in 174 students' researched composition papers from 16 disparate U.S. colleges and universities are working only with sentences from the sources and are drawing those sentences from pages 1 or 2 of the source 69 percent of the time, we can conclude that these papers offer scant evidence that the students can comprehend and make use of complex written text. Maybe they can; but they don't.

Our data raise the question of whether first-year students who are asked to write college-level researched papers have a full understanding of what that means. If they are told that their task is to make an argument and provide evidence supporting it from a number of sources, as Head and Eisenberg (2010) found many of our assignments require, then reading and engaging with those sources may seem counterproductive to the students. A reader who was sentence-mining this chapter might skip our methodology section entirely (indeed, in many disciplines this might be appropriate if the data are sufficiently clear); however, if that writer also skips the discussion, he or she might end up using our data as evidence for a claim that it cannot support.

Similarly, like several other authors in this collection (for example, see Purdy and Silva in Chapters 6 and 7, respectively), we do not present a thesis or finding until several pages into the chapter. A reader expecting a thesis on the first page might simply skip the entire chapter. Or, if challenged to summarize the argument in this chapter, an inexperienced reader of academic texts might report that we argue that writers "should be able to *talk* about the subject before [they] *write* about it" (a claim we quote from one of our sources on our first page). Another reader, having learned that we work on plagiarism, might search this document for terms such as "patchwriting" and use this article to provide a definition of that term or a statistic about its frequency, or maybe that reader would quote our recommendation that patchwriting be considered misuse of sources rather than plagiarism. Is any of that wrong? Not in the least. Would the reader have "digested" the broader argument? Not at all.

If writing instructors' goal in assigning the research paper is to use it as a vehicle to teach information literacy skills, synthesis of ideas, or argumentation, we seem to be failing. Our data, we believe, reveal a problem that our pedagogy should address. These and other Citation Project findings suggest a compelling need to overhaul the teaching of researched writing in college classes; what we are doing right now is producing results that no one can celebrate.

We hope that our campus librarians and our faculty colleagues in writing programs and across the disciplines will take these findings as a mandate for instructional change. For example, we believe that we must offer instruction designed to bring students to a deep engagement with sources, of the sort that enables them to talk *with* and *about* a source rather than merely mine sentences from it. This involves walking students through texts and modeling for them the kind of engaged reading and rereading that we expect of them. It also involves teaching and assigning summary-writing and the process of building summaries into a text. As Head and Eisenberg (2010) recommend, it means providing careful instructions for the researched paper that focus on the purpose and method rather than the punishment for failure to correctly cite sources. This research has led us as teachers to replace the end-of-semester researched paper with shorter papers that are source-based, but that use fewer sources and require students to engage with their arguments and build them into a conversation. At the very least, we urge our colleagues to focus attention not on the ethics of plagiarism, but on source use as "a sign of good workmanship, part of the morality of writing" as Perrin (1959, 636) puts it.

Endnotes

1. While the two of us, as principal researchers, have shepherded the work described in this article, many able, dedicated compositionists have worked as our co-researchers and are listed at www.citationproject.net (2012).
2. "Patchwriting" stands between quotation and paraphrase; it is neither an exact copying nor a complete restatement, and scholars such as Howard (1992) and

Pecorari (2003) have argued that it typically results from an incomplete comprehension of the source.

3. Examples of this include research on student information literacy skills by members of the library sciences and second language studies communities, and research on source use (and misuse) by psychologists and anthropologists.
4. Linda Smith (1981) elegantly describes what this type of research accomplishes: "In general, a citation implies a relationship between a part or the whole of the cited document and a part or the whole of the citing document. Citation analysis is that area of bibliometrics which deals with the study of these relationships" (83). See also Howard White (2004).
5. We give special thanks to Drew University Professor of Statistics Sarah Abramowitz, who generously advised us in this process.
6. We wish to thank Drew University for two faculty research grants, the McGraw-Hill corporation for an additional research grant to support the coding of data, and Binghamton and Syracuse Universities for providing staff and material support.
7. Like Mary Ann Gillette and Carol Videon (1998), we found tracking down these sources to be a challenge. In some cases we had to go through 30 papers to get 10 whose sources we could locate. That process taught us a lot about how much students struggle to identify the components of sources gathered electronically: Who is the author? What is the title? Who is the publisher? These things are far from clear to the majority of students whose papers we source-searched. But not all of the problems with source retrieval were because the student was at fault. Some institutions make available to their students collections of sources in databases such as the Opposing Viewpoints Series, to which our coders did not have access. This aspect of source selection is another finding of this research that we will explore elsewhere.
8. We have made our methods and training materials available to help people understand our data. The reliability and validity of Citation Project data comes from a methodology developed over half a decade and from careful training and calibration of coders. We believe that citation analysis can be a valuable pedagogical tool, a very effective part of faculty development, and a useful component in course and program assessment as we discuss at the end of this chapter. We do not, though, invite people to use our methods and identify them as Citation Project research without our permission.
9. Statistical Package for the Social Sciences (SPSS)—renamed Predictive Analytical Soft Ware Statistics (PASW), but still generally referred to as SPSS—is a series of integrated computer programs that allow researchers to record and review data and produce various forms of statistical analysis and reports. Tables 5.3, 5.4, and 5.5 in this chapter were generated by SPSS using the data we entered. Although PASW (formerly SPSS) includes a mechanism to test for inter-coder reliability

and variation among coder's decisions, we only entered final data once coding pairs had reconciled their coding sheets. For this reason we do not have PASW inter-coder reliability data. Because this research requires human judgment and interpretation, it is essential for coders to reach consensus on each individual citation. Where there were disagreements, one of the principle researchers joined the conversation to ensure consistency. The data for calibration papers coded by all coders therefore show 100 percent agreement rather than capturing the nuance of that conversation.

10. The Halo effect in empirical research, first described by Edward Thorndike in 1920 (25), occurs when one trait (in his case, physical attractiveness; in our case, effective writing) influences researchers' assessment of other traits (in his case, character; in our case, use of sources). More recent studies confirm his finding and add that the effect "extends to alteration of judgments about attributes for which we generally assume we are capable of rendering independent assessments," including in one example, students' writing (Nisbett and Wilson 1977, 250, 251).
11. For those unfamiliar with SPSS output tables, figures listed under "Valid Percent" are the percentages excluding any missing data. If any citations had been counted but not coded, that count would have been recorded in "Frequency" along with a percentage under "Percent," with the adjusted percentage of the five relevant traits appearing in "Valid Percent." In this case, all incidences of source use were counted and coded as one of the five traits, so "Percent" and "Valid Percent" are the same.
12. See the Council of Writing Program Administrators' Best Practices document for the differences between *plagiarism* and *misuse of sources* (www.wpacouncil.org/node/9). We agree that examples such as those presented in Table 5.2 should be defined as a misuse of source material, as should examples where the student omits to block or otherwise mark a cited quotation.

References

Anson, Chris M. 2008. "The Intelligent Design of Writing Programs: Reliance on Belief or a Future of Evidence?" *WPA: Writing Program Administration* 31 (3): 11–38.

Brown, Ann L., and Jeanne D. Day. 1983. "Macrorules for Summarizing Texts: The Development of Expertise." *Journal of Verbal Learning and Verbal Behavior* 22: 1–14.

Buranen, Lise, and Denise Stephenson. 2008. "Collaborative Authorship in the Sciences: Anti-Ownership and Citation Practices in Chemistry and Biology." In *Who Owns This Text? Plagiarism, Authorship, and Disciplinary Cultures*, edited by Carol Peterson Haviland and Joan Mullin, 49–79. Logan, UT: Utah State University Press.

The Citation Project. 2012. Accessed September 17, 2012. www.citationproject.net.

Corbett, Edward P. J. 1971. "The Theory and Practice of Imitation in Classical Rhetoric." *College Composition and Communication* 22: 243–250.

Gillette, Mary Ann, and Carol Videon. 1998. "Seeking Quality on the Internet: A Case Study of Composition Students' Works Cited." *Teaching in English in the Two-Year College* 26 (2): 189–194.

Head, Alison J., and Michael B. Eisenberg. 2010. "Assigning Inquiry: How Handouts for Research Assignments Guide Today's College Students." *Project Information Literacy Progress Report*. Accessed September 17, 2012. www.projectinfolit.org/pdfs/PIL_Handout_Study_finalvJuly_2010.pdf.

Howard, Rebecca Moore. 1992. "A Plagiarism *Pentimento*." *Journal of Teaching Writing* 11 (2): 233–246.

Howard, Rebecca Moore, Tricia Serviss, and Tanya K. Rodrigue. 2010. "Writing from Sources, Writing from Sentences." *Writing and Pedagogy* 2 (2): 177–192.

Maas, David. 2002. "Make Your Paraphrasing Plagiarism-Proof with a Coat of E-Prime." *et Cetera* 59 (2): 196–205.

Nisbett, Richard E., and Timothy D. Wilson. 1997. "The Halo Effect: Evidence for Unconscious Alteration of Judgments." *Journal of Personality and Social Psychology* 35 (4): 250–256.

Pecorari, Diane. 2003. "Good and Original: Plagiarism and Patchwriting in Academic Second Language Writing." *Journal of Second Language Writing* 12: 317–345.

Perrin, Porter, with Karl W. Dykema. 1959. *Writer's Guide and Index to English*, 3rd ed. Chicago: Scott Foresman.

Rosenwasser, David, and Jill Stephen. 2006. *Writing Analytically*, 4th ed. Boston: Thomson.

Smith, Linda. 1981. "Citation Analysis." *Library Trends* (Summer): 83–106.

Thorndike, Edward L. 1920. "A Constant Error in Psychological Rating." *Journal of Applied Psychology* 4 (1): 25–29.

White, Howard D. 2004. "Citation Analysis and Discourse Analysis Revisited." *Applied Linguistics* 25 (1): 89–116.