

Chapter 1

THE EVOLUTION OF THE CITATION PROJECT

Developing a Pilot Study from Local to Translocal

Sandra Jamieson

ABSTRACT

The historical narrative in this chapter traces the evolution of the Citation Project from its origins in a graduate seminar to the publication of pilot data (Howard, Serviss, and Rodrigue 2010) and the development of a transcontextual, multisite research project with internationally reported and replicated data. Based on interviews with principal and participating researchers and coders, analysis of research and coding notebooks, two blogs and various shared Google Docs, and e-mails as well as shared personal experiences, this chapter offers a historical account of methodological development that reveals the complexity and messiness of multisite research as well as the necessary adjustments that allow pilot research to be scaled to multisite projects. By being willing to expose not only their methods but also the false starts, challenges, and lessons they learned, Citation Project researchers hope to ease the transition to data-driven research and thereby increase the frequency of information-based policies and pedagogies.

INTRODUCTION

This chapter provides an antidote to the (necessarily) highly systematized accounts of research processes to which new researchers frequently turn, accounts that in users' minds too easily become ideals to be achieved and standards by which to measure their work. Books such as Johnny Saldaña's (2013), Stefan Titscher et al.'s (2000), and John Creswell's (2014) are invaluable procedural guides for conducting

research—and highly recommended—but while they do acknowledge the unruliness of qualitative research, they nevertheless present a linear, cleaned-up version of the process that can leave new researchers at a loss when their own work is stalled. Along with recent calls for writing studies researchers to share their methods and research design (Lunsford 2013), there is also a need for transparency in our field's research narratives. The reality of research, especially data-driven research, is that it is often a very messy, start-and-stop, revise-and-start-over process marked by frustration at many points along the way, as Rebecca Moore Howard and I noted in a keynote to the CCCC Research Network Forum (Howard and Jamieson 2012). Those of us trained in literary or rhetorical research methods are generally ill prepared for the challenges and time-consuming nature of data-driven research, and because it has not been a staple of our field until very recently, many of us lack mentors who can help. Similarly, most of us are unused to working collaboratively on research and writing, something probably essential for larger-scale research as our colleagues in the social and natural sciences learned long ago. There are many things to consider before beginning a RAD research project; this chapter presents some of those factors in hopes of encouraging other such endeavors.

Collaborative RAD research is infinitely more rewarding than anyone imagines, though, and, as the other chapters in this book reveal, has the potential to lead to the kinds of changes in pedagogies, policies, and practices many of us desire. I believe research narratives that are honest about failures and setbacks, coupled with the methods and design of the final research projects they engendered, will help researchers—experienced and prospective alike—imagine and plan large-scale research projects of their own. I hope narratives like this one will also help my fellow researchers work through the inevitable messiness and rethinking that brings such projects to successful completion.

The research project that is the focus of this chapter is the Citation Project, specifically a study of eight hundred pages of researched writing produced by 174 students enrolled in first-year writing courses at sixteen US colleges and universities. Researchers coded both the kinds of sources selected and the ways students incorporated information from those sources into their papers (summary, paraphrase, quotation, patch-writing, or copying). They also coded the kinds of sources used, including type, length, and reading difficulty.

The methods and findings of the Citation project sixteen-school study have been described elsewhere (Jamieson 2013; Jamieson and Howard 2013), and documents from that research are included in the appendix

to this chapter.¹ My purpose here is not to describe those methods *per se* or discuss the findings (although I will mention them by way of comparison) but to narrate the evolution of the project's procedures and coding methods over a considerable time and through a series of messy drafts that ultimately allowed the collection and analysis of transcontextual RAD data on a broad scale. Using information from interviews with founding researchers (principal and participating researchers and coders), analysis of research and coding notebooks, two blogs, and various shared Google Docs and e-mails, in addition to personal experience, I will describe the various challenges encountered as the research moved from a series of questions generated in a graduate seminar to a single-institution study, then to a three-school study conducted after I became one of the two principal researchers, and thence to the sixteen-school study whose data I reported above. By sharing not only their methods but also the false starts, challenges, and lessons they learned, Citation Project researchers hope to ease the transition to data-driven research and thereby increase the frequency of information-based policies and pedagogies.

While there are things we all wish we knew before we started this research, what is more useful to future researchers is what we learned in the process and the ways it led us to refine our research and develop methods we can share with others who are conducting their own citation context research or replicating the sixteen-school study.

CITATION PROJECT ORIGINS

It is instructive to trace the development of large-scale research like this to the various points of origin, both to give credit to the many people involved and to emphasize the importance of ideological and theoretical frameworks that necessarily shape research. Going back to the original motives and influences can help researchers when they become blocked, reminding them of the reasons they are doing what they are doing and how many problems they have already overcome—which is why the kind of record keeping that led to this chapter is so important.

Origin Stories

Everything has an origin story; the Citation Project has two.

Narrative 1: The Linear Narrative

Rebecca Moore Howard first became interested in student source use in the mid-1980s, coining the term *patchwriting* in 1993 when she described

her analysis of writing by students in one of her classes (Howard 1993). A decade later, Diane Pecorari set out to empirically test Howard's claims in the writing of second-language graduate students (Pecorari 2003). These two articles helped shape a doctoral seminar in curriculum design focused on authorship studies at Syracuse University in the fall of 2006, from which developed a small class project that experimented with textual coding and then developed into an actual pilot study at a single institution. The results of that study were published in 2010 (Howard, Serviss, and Rodrigue).

I joined the project in 2008. While Serviss and Rodrigue turned their attention to other projects, Howard and I expanded the single-institution pilot study to three institutions (a liberal arts college, a private research university, and a state university). We brought in contributing researchers to help test the citation-analysis methods and code the papers, and we described our findings in a presentation entitled "The Citation Project Three-School Study" at the Conference on College Composition and Communication (Benedicks et al. 2010; Jamieson 2010).

After that three-school study came a plan to collect and code papers from ten colleges and universities that would represent a wide geographic distribution and institutional variety. Ultimately, 174 papers from sixteen institutions were collected and coded and the initial data from all sixteen was presented at the CCCC conference two years later (Jamieson 2012). Although Citation Project researchers generally refer to the initial single-institution study as the pilot study in that it developed the general coding categories used to code source use throughout the expansion of the project, the coding procedures and terminology continued to be refined in the process of conducting the study of papers from three quite different institutions, and those papers were ultimately recoded using the final criteria developed for the sixteen-school study. New research is now being conducted by scholars involved in that initial study, and new projects are developing in the United States and abroad, necessitating more fine tuning of methods to accommodate different citation styles. In 2012, principal researchers Rebecca Moore Howard and Sandra Jamieson were joined by Tricia Serviss, a co-researcher for the single-school pilot study and coauthor of this collection. Serviss and I are also working with Angela Feekery to revise and repeat the study across universities in her native New Zealand.

Narrative 2: Theoretical Underpinnings

There is also a second origin story. This one concerns not the Citation Project research per se but the theoretical and methodological

frameworks that made it not only possible for Howard and I to imagine doing such research but impossible for us not to do so. And the origin story here is the moment when we were confronted by the realization that our field needs data-driven research and we needed to do it. This realization led me to take a statistics class at my institution and therefore to be able to talk about the language of statistical analysis when I read a draft in progress of the article “Writing from Sources” (Howard, Serviss, and Rodrigue 2010); it also led me ultimately to become a principal researcher with the Citation Project.

The realization about the importance of data-driven research came as we listened to Chris Anson’s keynote presentation at the 2006 conference of the Council of Writing Program Administrators (Anson 2006; 2008). In that keynote, Anson argued that the shared assumptions of writing studies researchers and practitioners make it relatively easy to make a case for change others would endorse. Outside our own field, though, those shared assumptions do not necessarily prevail, and the *ethos* of the person making the argument is a much less powerful piece of evidence than it is within our field. Anson contended that if writing program administrators are to persuade cross-curricular colleagues, higher administration, fund-granting foundations, legislators, and the like to do or change anything, those WPAs need data-based evidence. Statistics, he specified, are the gold standard of universal evidence.

Rebecca Moore Howard and I were sitting side by side in the audience in Chattanooga when he delivered that keynote, and we spent the remainder of the conference talking about its possible implications for our own work. We two had been coauthors since 1993, when we began *The Bedford Guide to Teaching Writing in the Disciplines* (Howard and Jamieson 1995), and we shared scholarly and pedagogical interests. This collaborative history is important because it allowed us to work through the inevitable disagreements and setbacks as we developed and worked on the Citation Project. However, what has really motivated us through this work and continues to motivate us as we write and speak about the data and expand the project is the idea that statistical and transcontextual evidence can bring change in ways anecdotal evidence cannot (see Howard 2011, 2014). Before Howard published her data about the students in her class in 1993, other scholars had published individual case studies highlighting the challenges experienced by developmental writers as they try to incorporate sources (Hull and Rose 1989 and 1990, for example) and try to build papers from those sources (Kantz 1990; Kennedy 1985), but the *numbers* provided in Howard’s 1993 article had resonance and, of course, she also named the phenomenon.

Once named and defined, patchwriting could be identified as such and measured, and scholars could begin the process of establishing it as a separate category from plagiarism (Jamieson 2016). The frequency with which the 2010 article describing Howard, Serviss, and Rodrigue's initial study is cited speaks to a larger thirst for data-driven findings. It is the need for more data that motivated Howard and me because we both want to see the pedagogical and policy changes Anson argues, and we agree, only data can bring.

DESIGNING CITATION CONTEXT RESEARCH

Howard shared a copy of Anson's keynote with students in her doctoral seminar in fall 2006 along with the recommendations of the CCCC Caucus on Intellectual Property. In response, the students—Sarah Etlinger, Tanya Rodrigue, Tricia Serviss, Zosha Stuckey, and Terri White—set about exploring how such data-driven research could be used to investigate authorship issues (the topic of the class). They focused on how to study what undergraduates do with the sources they use in their college papers and collectively designed (and named) the research project described in the class blog (Figure 1.1) and drafted information for an IRB application.

The “pre-determined units of analysis” (Figure 1.1) were those developed by Diane Pecorari (2003)—transparent and opaque source use—to test Howard's hypotheses from ten years earlier (Howard 1993). Working with nonnative speakers of English, Pecorari took ten randomly selected pages from portions of seventeen draft dissertations and collected an additional ten pages from each of eight published PhD theses. Samples were “divided into passages of varying length, the passage boundaries being determined by the source use” (Pecorari 2003, 322), allowing her to compare each passage with the source from which it drew. She coded source use as *transparent* or *opaque*. If one could tell where a source was used (ideally where it began and ended), the citation was transparent; where it was difficult to separate the source from the student prose, the material was coded as opaque. Working from this same method of citation context analysis, the seminar participants set out to replicate Pecorari's research by coding five student papers from their own institution. While no coded material was preserved, the blog contains discussions of individual sections and the struggles the researchers experienced as they tried to code (“CCR732F06” 2006). Without definition, the codes were difficult to consistently apply and the coders simply couldn't find consensus. By

Project Grande v. 2.0 – Operation Writer Freedom!

How do students use sources in their academic writing? Why do students use sources in their academic writing? How do students draw on sources for information? How do students understand the purpose and ethical use of sources in academic writing?

These questions stem from the recommendations of the CCC -Intellectual Property Caucus to compositionists, urging their involvement in “leadership roles...educating their institutions about the limitations of [plagiarism detection services] and conduct[ing] more empirical research to understand better how these technological services affect student’s writing and the educational environment.” In the spirit of their counsel we have designed a study of student texts which will allow us to examine undergraduates’ composition choices as they relate to and complicate ideas of source use and plagiarism. Our project - **Operation Writer Freedom** - will be initiated in the Spring of 2007 at Syracuse University as a pilot study with the projected intention of developing into a national endeavor on a variety of campuses. Ultimately, our goal is to answer CCCC-IP’s call for responsible research which will help promote productive pedagogy and aid in a better comprehension of student work when engaging sources.

The Process of Operation Writer Freedom

- * We will ask willing instructors to distribute release forms to students, allowing us, through IRB, to use undergraduate texts for this study.

- * We will collect course syllabi and text submissions from participating 205 pupils.

- * Students will electronically submit their papers to a centralized digital location at the same time as they submit finished work to instructors.

- * **Operation Writer Freedom** team members will code these texts using pre-determined units of analysis, carry out a trial comparison of the texts utilizing diverse textual analyses’ methods, store the texts, and process the data in order to learn more about student source use.

Figure 1.1. Class blog post from November 2006 showing draft IRB application for a Citation Project precursor

the time they had tried to code three papers, they had rejected the terms *transparent* and *opaque* as insufficient. They also concluded that to understand how the sources were being used, one would need to read every cited source, something they had been unable to do and, they worried, an unsustainable method. So here in this early modification of both coding method and practice are two important realizations that helped shape Citation Project research: *all cited sources must be read by coders; all coding categories must be clearly defined.*

The goal defined in this blog post—to develop “responsible research which will help promote productive pedagogy and aid in a better comprehension of student work when engaging sources”—remains the goal of the Citation Project today, and the process they describe can also be traced in current methods. Ultimately, although the class did not complete the research they designed, Citation Project research did follow the trajectory described in Figure 1.1 and “develop[ed] into a national endeavor on a variety of campuses.” The process of moving from initial

idea to action was not the smooth linear process many people imagine when they set out to do research—or when they hear single-point origin stories. Many theoretical discussions, microstudies, revisions, and collaborations go into the design and execution of successful transcontextual research, and frequently people leave and others join along the way. Such fluidity is alien to writing studies research but common in other fields where transcontextual research is the norm.

FROM DESIGN TO PILOT STUDY

The conclusion of the seminar participants that researchers needed to move from overly narrow coding categories to broader and more focused categories is typical in the development of coding projects (Saldaña 2013, 11). It is not surprising, then, that at the conclusion of the semester when Howard, Serviss, and Rodrigue decided to revise the initial ideas explored in the seminar into a pilot study, their first step was to work on coding categories. The three researchers spent a semester designing coding categories, developing the research plan, securing IRB clearance at their institution, and gathering student papers produced in a required research-writing course. In this first pilot phase, Howard, Serviss, and Rodrigue coded eighteen student papers that had been submitted for a grade in fifteen sections of a composition course at a private, not-for-profit university with an RU/H Carnegie basic classification (Howard, Serviss, and Rodrigue 2010).

To do this, they still drew on Pecorari's citation-analysis methods, but they coded incidences of copying, quotation, paraphrase, patchwriting, and summary—focusing on explicit textual moves rather than reader perception of opacity and transparency. The three researchers had spent months developing methods of analyzing students' source use in the papers so they might reach consensus about their results, and a key aspect of that was articulating clear definitions of each coding category. They defined *summary* as “restating and compressing the main points of a paragraph or more of text in fresh language and reducing the summarized passage by at least 50 percent” (Howard, Serviss, and Rodrigue, 2010, 181), and they found no instances of it in any of the papers they coded, even with summary referring to a source passage as brief as a paragraph. Instead, the coding revealed a great deal of *patchwriting*, which they defined in the pilot as “reproducing source language with some words deleted or added, some grammatical structures altered, or some synonyms used” (181). In fact, patchwriting was found in sixteen of the eighteen papers (182), or 89 percent. In this pilot phase,

researchers coded for presence or absence of textual behaviors, not for frequency. That approach was to change as the research expanded. The definitions of the key terms evolved as well.

Howard's 1986 introduction to the phenomenon of patchwriting reports that nine of the twenty-six first-year college students in her general education course—34 percent—patchwrote at least once in their papers (Howard 1993, 237). In Pecorari's study of seventeen nonnative speakers of English enrolled in graduate programs at UK universities, sixteen of the seventeen—94 percent—"had one or more passages in their writing samples in which 50 percent or more of the words came from their sources without being indicated as quotation" (Pecorari 2003, 325). Pecorari's data are in line with the 89 percent finding by Howard, Serviss, and Rodrigue (2010, 182). Their study, then, shed light on how many first-year students might be patchwriting in their researched papers, expanding the focus of earlier studies (Hull and Rose 1989 and 1990, for example) beyond "developmental writers" to the general student population. With patchwriting widely reviled among college instructors and described as academic dishonesty in many college plagiarism policies and textbooks, it was important to know that sixteen out of eighteen students included some of it in the researched papers they wrote for their college composition class. The research followed Howard's (1993) interpretation of patchwriting as an example of "a summary technique characteristic of writers in difficulty, or writers in relatively early stages of cognitive development" (Howard 1993, 237). Rather than an act of academic dishonesty, patchwriting was "a healthy effort to gain membership in a new culture" (236) by doing what David Bartholomae (1986) had called "trying on the discourse" of that culture (6), an interpretation that has now become widespread.

This pilot study offered the first data from general first-year composition classes, and it contributed additional evidence to support the increasingly inarguable claim that most school writers—perhaps at all levels, perhaps in all classes—patchwrite sometimes as they work from sources. All three studies (Howard 1993; Howard, Serviss, and Rodrigue 2010; and Pecorari 2003) involved naturalistic data produced by school writers. With Miguel Roig's experimental research revealing that 22 percent of psychology professors patchwrote when trying to summarize a difficult text from outside their field (Roig 2001, 315) and 68 percent of students did so when asked to reproduce the ideas in one single paragraph (Roig 1999, 976), it was plausible to assert that patchwriting is commonplace in writing from sources and therefore requires a pedagogical rather than a punitive response.

WHY REPLICATE THE STUDY?

The other major finding of the study—that students work from sentences in the source rather than extended passages (Howard, Serviss, and Rodrigue 2010, 186)—has held true through all subsequent studies. Similarly, while the national study found patchwriting in only 52 percent of the coded extracts, the finding that it is a common phenomenon that occurs with much greater frequency than previously believed was confirmed. For many writing studies researchers, this confirmation might suggest it was unnecessary to do any additional research after the first study; however, it took replication in the national study to confirm the findings of the pilot. Without that replication, the pilot was just another single-school study and it was impossible to know whether the findings were broadly representative of student source-use practices or simply a result of the local institutional context of the student writers studied. For data to have traction, it must be representative and therefore transcontextual, as Anson (2006) proclaimed and the course blog post noted (Figure 1.1).

*RAD Research: The Relationship between Pilot
and Sixteen-School Study Findings*

While the full findings of the sixteen-school study are available elsewhere (CitationProject.net; Jamieson and Howard 2013), it is useful to include two sets of data here for comparison. The first (table 1.1) shows the frequency with which each kind of source-integration method occurred within the 1,911 citations. The second (table 1.2) shows how many coded extracts included at least one incidence of each form of source integration.

While the pilot study found no summary, the sixteen-school study found that 6 percent of the citations were summary (table 1.1) but that 41 percent of the papers included at least one instance of summary (table 1.2). This finding suggests not so much that students need to learn how to summarize, as the pilot concluded, but that they need to learn how to summarize more frequently, digesting ideas and incorporating them into papers instead of working from sentences in the source. Similarly, the smaller percent of coded extracts that include at least one instance of patchwriting—89 percent in the pilot and 52 percent in the national study (table 1.2)—is further complicated by the fact that 78 percent of the coded extracts include at least one incidence of paraphrase and all but three of the extracts that include patchwriting also include at least one instance of paraphrase.

Table 1.1. Frequency of source integration method in 1,911 student citations

<i>Predominant Use of Source Material within the Citation</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Percent</i>
Copying cited but not marked as quotation	83	4.34	4.34
Copying cited and marked as quotation	793	41.50	45.84
Patchwriting	306	16.01	61.85
Paraphrasing	609	31.87	93.72
Summarizing	120	6.28	100.00
Total	1,911	100.00	

Table 1.2. Type of source-integration method occurring at least once per five-page extract from 174 papers

	<i>Frequency</i>			<i>Percent</i>		
	<i>Occurs at least once</i>	<i>Does not occur</i>	<i>Total</i>	<i>Occurs at least once</i>	<i>Does not occur</i>	<i>Total</i>
Copying not marked as quotation	33	141	174	18.97	81.03	100
Copying marked as quotation	159	15	174	91.38	8.62	100
Patchwriting	91	83	174	52.30	47.70	100
Paraphrasing	135	39	174	77.59	22.41	100
Summarizing	71	103	174	40.80	59.20	100

These data reinforce the finding of the pilot study that the phenomenon of patchwriting is more frequent than we think but also strengthen the case that cited patchwriting is simply poor writing in need of revision. Coded extracts reveal students slipping from paraphrase to patchwriting and then back to paraphrase—often using the same source (see Figure 1.3)—indicating they were probably unaware they had misused the source. These data, then, suggest not that we must rush to purchase plagiarism-detection software (PDS) but that we must teach students how to revise patchwriting into paraphrase or summary.

By building on the methods and goals of the initial conceptualization of the study in a doctoral seminar and then replicating key aspects of the pilot research, the full study was able to confirm the basic findings of the earlier work, but it also deepened our understanding of the ways writers

incorporate source material into their prose. In addition, the sixteen-school study expanded the research to include analysis of the sources themselves, replicating the categories used in single-site studies by Carlson (2006) and McClure and Clink (2009) and allowing researchers to trace correlations between source type and source use and to explore sources and information-literacy skills more deeply (see Jamieson 2017).

FROM PILOT STUDY TO TRANSCONTEXTUAL RESEARCH

As Serviss and Rodrigue turned their attention to their dissertation projects, Howard and I set about developing a method to extend the research beyond the single institution, as the class blog had proposed (Figure 1.1).

Developing Coding Categories

The first step, as it had been for Howard, Serviss, and Rodrigue, was to explore what features should—and could—be coded. At first the coding sheet became more and more complicated as we listed more and more questions we would like to be able to answer. My research notebook reminds me that by the end of a meeting in June 2008, there were six categories, each subdivided into at least two subcategories. If that coding sheet had been used, researchers would have been coding naturalistically produced text for sixteen separate features of source integration, some that included embedded subcategories such as “misrepresents what the source is saying on a denotative level or attributes something to the source that it doesn’t actually say” and “copies from a source, makes copying errors, marks it as a quotation, and cites the source.” The papers were to be coded electronically, with coding sheets assigning a different color to each of the sixteen categories. Coders were to highlight each category with one of these colors. When they had coded the paper, they were to total the number of times they had marked each of the sixteen categories and enter those totals into the form, giving a frequency count for each.

Part of the reason for this added complexity was my arrival on the scene having not participated in the previous studies. Howard, Serviss, and Rodrigue had spent months developing their coding categories, but it is common to revisit categories when a new researcher joins the team. I had not yet coded and had no idea how impossible my enthusiasm for data would be. We got increasingly excited about the many things our data set could reveal about student source engagement and so developed more and more increasingly complex categories. Finally, we realized that less is more and we did not need to code everything at

once, a lesson we consider essential for code-based research. We began reducing and ultimately returned to the five categories of the pilot study (copying but not marked as quotation, copying marked as quotation, summary, paraphrase, and patchwriting). The sources selected and used were coded separately by a separate team of coders who combined the data into the final SPSS document (see Jamieson 2017).

While Howard, Serviss, and Rodrigue had done their own coding, expanding the project also meant expanding the number of people who would code the papers and bringing in coders unfamiliar with the project and the long conversations about categories and definitions. As training began, we realized the importance of category descriptions. On a research trip to the United States in 2009, experienced coder and plagiarism researcher Wendy Sutherland-Smith took part in a two-day coding session of papers from one of the first three institutions. The coding categories had already been streamlined somewhat by then, but still the coders struggled to apply them to the papers, and Sutherland-Smith gently suggested that the problem was the categories themselves. She was right.

The category causing the problem was *patchwriting*, a word that carries negative connotations for many people. We began with Howard's 1993 definition ("copying from a source text and then deleting some words, altering grammatical structures, or plugging in one-for-one synonym substitutes," 233). That definition had changed pedagogy and policies and led to other scholarship; however, we soon found that a *definition* of patchwriting was not necessarily helpful to coders who came to the task with their own well-established definitions of patchwriting as plagiarism. The use of language in coding terms that has negative associations predisposes coders who identify an instance of that feature in the *text* to view the *author* of that text negatively (as *unethical* in this case) and so to transfer that negative predisposition back to the remainder of the paper they are coding. This phenomenon, in which an association with one feature produces a similar association with the whole, is known in psychology as the *halo effect* (Nisbett and Wilson 1977; Thorndike 1920). Once the association is made, coders are not only more inclined to find examples of the negative feature but also to actively look for it. Terms and phrases like *copying*, *deleting*, *altering*, and *plugging in one-for-one synonym substitutes* suggest patchwriting is a deliberate attempt to conceal borrowing and implicitly prompt coders to assess the intention of the author rather than simply describing what they see in the text. This increases the likelihood of miscoding. In early coding sessions in particular, once they had identified a few incidences of patchwriting, coders often identified quite clear examples of paraphrase as patchwriting

further into the paper. Understanding the halo effect led us to suspect that once they had identified an incidence of patchwriting, coders were inclined to think of the author of the paper as “a patchwriter” and as a result to be on the lookout for more examples and to see them where they weren’t.

We also found a second cause of miscoding. This one was connected to the quality of the student prose in general and our tendency to associate patchwriting and plagiarism with weak writers. When a paper is deemed to be well written overall, patchwriting tends to be initially coded as paraphrase, or even summary; when the overall prose, organization, or grammar of the paper is weak, paraphrase is frequently coded as patchwriting. This, too, is an example of the halo effect, and it may also explain why faculty more typically find patchwriting in the work of weak writers but not in that of stronger writers despite the Citation Project’s finding that more than 50 percent of all writers in the sample patchwrote at least once.

Coders needed a *description* that would focus attention on what they saw in the text and so help them agree upon where they saw instances of patchwriting and how they could distinguish it from paraphrase or copying. But they also needed a description of patchwriting that forced them to ignore the connotations of the word and placed the *text* in the subject position rather than the writer. We were reminded by those who have studied the halo effect that a category must help coders “report the evidence, not a rating” (Thorndike 1920, 29). By the time we led a preconvention workshop at the 2010 CCCC, we were using a description of patchwriting in which a passage of text is described as “restating a phrase, clause, or one or more sentences while staying close to the language of the source” (Benedicks et al. 2010). This description neither speculates on nor insinuates why or how the students stayed close to the language of the source; it just describes the fact that they do. And it is sufficiently reliable to allow us to achieve acceptable levels of inter-rater reliability. Coders were instructed to use this definition and remain descriptive in their coding of all features; coders ought to ask *what is happening in the text in front of me? How is the source material being incorporated?*

Responding to Economies of Scale

What to Code

We had already collected papers from two schools in addition to the institution in the pilot study, and in fall 2009 and spring 2010, we collected papers from thirteen more institutions. It quickly became apparent that

with institutions contributing papers whose lengths varied considerably, data from some of those schools would skew our results. After consulting with several statisticians, we decided to code the same number of pages from each of ten papers from each institution. We selected pages 2–6 because there is less source use on the first page of researched papers and because most of the student papers sent to us were at least seven pages long. This decision to code partial papers allowed us to compare institutions without having to adjust for sample size, and an assessment of patterns within the entire sample would give us the desired “snapshot” of methods of source integration across contexts. *This was the second of several revisions that made it possible for us to move from a small study at one institution to a large study drawing on different sites, one that ultimately allowed us to code eight hundred pages of student prose.*

In January 2010, we began to analyze ten papers from those two additional institutions and to recode ten randomly selected papers from the pilot institution. At that point, we were coding every sentence of entire papers, an extremely laborious process even with only ten papers from each institution. Other aspects of our method were also still evolving as we worked with these papers, and in many ways the three-school study functioned as a more advanced pilot for the sixteen-school study (indeed, once the coding methods and categories were finally settled, we had to recode the paper extracts from those first three institutions). When we began this stage of the research, coders identified source blocks as they read, and when they found material that appeared to come from a source, they read the source and coded how it was used. If the material did not appear to come from the source cited, they searched all the sources in the Works Cited list (a process made easier because the sources had been saved as searchable PDF documents, but still long and arduous). Sometimes they found that the wrong source had been cited, so they coded the type of source integration and noted on the coding sheet and in the margin that the wrong source had been cited. While it was agreed that only material from cited sources would be coded, we composition instructors have been so heavily inscribed by an ideology of “seek out all plagiarism” that it was sometimes too difficult for coders to ignore uncited material that appeared “obviously” to have been drawn from a source. My research notebook describes a memorable coding session at which two of the coders spent several hours each on an exhaustive (and unsuccessful) web search for material they believed to have been plagiarized—even though our coding guidelines clearly excluded such searches from the process.

What Not to Code

That detour—and the need to avoid others like it if coding was to be completed on any reasonable schedule—led to a third important procedural revision, one that more than any other made it possible to scale up the research from a localized pilot to a multisite, multicoder study and that also shaped the way we described the process. My research notebook records the decision this way: “The Citation Project is concerned with the ways students USE material from the sources they cite. If material is not cited, we ignore it. If it is incorrectly cited, we mark it as such and then move on.” The notebook heading for this decision is “No Hunting,” with this additional observation in red ink: “*This research is not about [catching] plagiarism; it is about what happens when students use sources that they cite.*” We revised our method to prevent plagiarism hunting and focus coders’ attention on the text the student cites as coming from a source. Before coders begin coding a paper, the researcher responsible for finding the sources marks up the paper, drawing a box around material indicated as cited and noting the alpha-numeric code assigned to the cited source (see Figure 1.2). The coders then *only* code what is in those boxes. If the source cannot be found, or if it is the wrong source, the box is not coded, but the coder marks *source not found* or *incorrect source cited* in the margin of the paper and on the coding sheet (see the coding guidelines in Appendix 1.A).

The Need for Systemization

The fourth and final significant change came as we tried to recode paper extracts following the previous procedural revisions. Coders had been following the method used by Howard, Serviss, and Rodrigue in the pilot study: highlighting material from the source with whatever color highlighter came to hand and then writing in the margin what they found (see Figure 1.3). With many more coders, this process became confusing at best. Each coder had an individual style of marking up the paper, making it often impossible for a second or third reader to understand how specific coding decisions had been reached, and sometimes even what those decisions were. Ultimately, the solution drew on the long-abandoned idea of sixteen colors for sixteen categories: each different method of source integration (paraphrase, summary, patchwriting, and quotation) was to be highlighted with a predetermined color (green, pink, yellow, and blue, respectively). In order to facilitate the new color-coding system, we provided coders with highlighter pens adorned with stickers indicating the color categories and later a multicolored marker

<p>4 S0402 Cont.</p>	<p>theorists, the classics was not only a vast resource of universal ideas, but there was encouragement in the fact that Classic writers themselves advocated imitation as the best means of developing style, expression, and taste" (Buelow).</p>				
<p>5 S0401</p>	<p>In Penelope Alfrey's article, she describes that in literature and in the visual arts, from the Renaissance onwards, an abundant amount of models encouraged artists to engage in copying or imitation. Then there came the distinction between copying and imitation. Imitation has to have generalization to some degree, for this allows the imitator to move around and add things as they desire. Copying is defined as a mechanistic form of reproduction. Through copying an apprentice learned the necessary techniques and skills required of an assistant in a studio. Copying was also a commonly accepted practice as a means of reproduction. The practices of artists and designers, until the early 20th century, relied heavily on copying.</p>				
	<p>However, predictability, the thing people were used to getting, became offset by the unpredictable. Unpredictability became what people wanted, but unpredictability was seen as an asset that signified genius and therefore it was not an asset everyone had. From then on it was not good enough to just copy and imitate. They were expected to invent.</p>				
<p>6 S0401</p>	<table border="1"> <tr> <td data-bbox="264 920 688 1006"> <p>Alfrey believes that copying is integral to creativity, "With no opportunity to plagiarize, the imagination appears to suffer" (Alfrey).</p> </td><td data-bbox="688 920 966 1006"> <p>Historical evidence shows that</p> </td></tr> <tr> <td colspan="2" data-bbox="264 1006 966 1045"> <p>standards are greatly determined by precedents, the things that have come before.</p> </td></tr> </table>	<p>Alfrey believes that copying is integral to creativity, "With no opportunity to plagiarize, the imagination appears to suffer" (Alfrey).</p>	<p>Historical evidence shows that</p>	<p>standards are greatly determined by precedents, the things that have come before.</p>	
<p>Alfrey believes that copying is integral to creativity, "With no opportunity to plagiarize, the imagination appears to suffer" (Alfrey).</p>	<p>Historical evidence shows that</p>				
<p>standards are greatly determined by precedents, the things that have come before.</p>					

Figure 1.2. Extract with citations blocked for coding. Material not boxed is not coded.

with a label. Coders would also note in the margin what they had found and the page in the source from which it was drawn and transfer this information to coding sheets (see Appendix 1.B), allowing the principal researchers to double check and recode paper extracts easily, as the samples in Figures 1.3 and 1.4 demonstrate.

Over the life of the research, we relied on many coders (who are named on our website, CitatonProject.net), and our training procedures became more effective each time we ran training and group coding sessions. Eventually, each paper was read by two coders who coded the paper individually and then created one reconciled coding sheet (see Jamieson and Howard 2013). If agreement could not be reached, one of the principal researchers stepped in and recoded. We

marketing drugs. Ghostwritten journal articles always contain undisclosed conflicts of interest. Ghostwriting, in their eyes, undermines science, as well as fails to give the readers 100% reliable information (Elliot and Moffat).

Ghostwritten articles may mislead doctors about the actual risks and benefits associated with medical treatments. Medical misinformation of this sort has the risk of harming a large number of people. The potential for harm is amplified by the fact that ghostwritten articles usually bear the name of a highly respected researcher who appears to have no financial risk in the issue at hand (Elliot and Moffat). Erica Johnson gives her definition of medical ghostwriting as "The practice of drug companies drafting review articles endorsing new medicines and presenting them to prominent doctors and scientists to put their names to, in exchange for financial (or other) compensation". She also says that time and money are both factors for why ghostwriters are used, and overworked doctors have become willing to serve as "authors" for papers written for them by ghostwriters paid by drug companies. Ghostwriting can now be found in all major medical journals, as opposed to when it was only limited to journal supplements sponsored by drug companies. Some cases have shown that the doctors listed as authors did not review the raw data (tables compiled by company employees) that they were writing about.

Handwritten annotations on the right side of the text include:

- 4 S1002
- 5 S1002
- 6 No page #s are given for paginated sources
- 7
- 8
- 9
- 10
- S1002
- S1004

Figure 1.3. Paper from the three-school study coded using the pilot-study method. See color versions of this coding at CitationProject.net/codingexamples/.

also randomly recoded about 10 percent of the paper extracts to ensure continuity. Initially this method was used only to calibrate coders at the beginning of each session, after which they would code alone; however, we soon realized double coding and discussion led to much more accurate coding, so that is how the papers were coded.

<p>⑪ Z1511</p>	<p>discovery of the endocannabinoid regulatory system which is the hypothesis that, "humans, and all animals, make and use internally produced cannabis-like products (endocannabinoids) as part of the evolutionary harm reduction program" (Melamede).</p>	<p>Copy & Quote p.1</p>
<p>⑫ Z1513</p>	<p>Evidence of a jump in interest can be seen in a jump from 258 journal articles published in 1996 on cannabis, to over 2,100 published studies published in 2008 (Recent Research on Medical Marijuana).</p>	<p>Paraphrase p.2</p>
<p>⑬ Z1513</p>	<p>As a result, scientists are now looking at the potential for cannabinoids to be used to modify diseases. More specifically, scientists are looking at methods using cannabinoids to moderate autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, and inflammatory bowel disease, as well as their role in the treatment of neurological disorders such as Alzheimer's disease and Lou Gehrig's disease (Recent Research on Medicinal Marijuana).</p>	<p>Patchwritten p.2</p>
<p>⑭ Z1516</p>	<p>There have also been recent findings by doctors in the Netherlands that suggest THC helps to reduce neuronal injury through the endocannabinoid system, thus protecting the brain from neurodegeneration (Stelt). Most importantly, investigators are now studying the anti-cancer properties of cannabinoids.</p>	<p>Paraphrase (close) p.1 (from abstract)</p>
<p>⑮ Z1513</p>	<p>There is an increasing amount of preclinical and clinical data that conclude that cannabinoids stop the spreading of specific cancer cells through programmed cell death and the prevention of the forming of new blood vessels (Recent Research on Medicinal Marijuana). On top of all this, there are 19 major diseases that are undergoing clinical trials of the therapeutic effect of cannabinoids.</p>	<p>Patchwritten p.2</p>

Figure 1.4. Paper from the final study coded using the four-color, boxed-citation method. See color versions of this coding at CitationProject.net/codingexamples/.

Refining Methods

Working with multiple coders and a large sample, then, led to five significant changes to our method as we transitioned from pilot to trans-contextual research.

- We developed text-focused coding definitions.
- We decided to code only pages 2–6.
- We moved to coding only text in marked citation blocks.
- We asked coders to use predetermined color-coded highlighting.
- We always employed double coding, discussion, and then reconciling coding sheets.

The team successfully recoded paper extracts from the first three schools using this revised method. Had we not made these changes, it is unlikely the research would have been completed. While these modifications reduced much of the complexity initially imagined, these simplifications made it possible to code for one thing and to code a paper in about an hour. With two coders per paper, followed by a conversation,

the process was still time-consuming; however, the five or more hours it was taking each person to code a paper in the pilot study simply was not scalable for 174 papers.

As we continued random recoding, we realized some coders were sometimes wildly inaccurate in their coding, and on closer examination, we determined this to be the case when they were knowingly coding papers from their own institution. Although the papers were anonymous and not from any of their classes, we initially employed coders from the submitting institution at least for the first round of coding, but sometimes for both. A version of the halo effect seemed to be in play again: coders were more likely to be generous in their coding of patchwriting as paraphrase in papers they knew to be from their own institutions. So came a sixth adjustment.

- *We agreed coders should never knowingly code papers from their own institution.*

This means we encourage those who would conduct local citation context—or probably any other partially or fully subjective research—to pair with at least one other institution and code each other's papers.

Although we had simplified how coders assessed the method of integration of source material, we found ourselves intrigued by the sources themselves. I had also been working on information-literacy issues, so we agreed to also code the sources, both to gain a deeper knowledge of what kinds of sources were being used and to expand the many single-site studies whose findings were dominating the field (Carlson 2006; Davis 2002; Jenkins 2002; McClure and Clink 2009). We paid a separate set of coders to mark up the papers as described above, find the sources cited in the five pages selected and make PDF copies of them, and then add to a spreadsheet information about the source using a slight modification of categories developed by Carlson (2006) and McClure and Clink (2009). These source coders also ran each cited source through a program that assessed textual difficulty level using Flesch reading-ease and Flesch-Kincaid grade-level analysis (Flesh.app). With source information added to the spreadsheet for each citation, we were able to look for more nuanced correlations about how students incorporate material from different kinds of sources (books, journal articles, websites) and sources with different lengths and different difficulty levels (allowing us to test Roig's claims that writers more frequently patchwrite from difficult sources). By having different coders working on different parts of the project, we avoided confusion and developed richer data (see Jamieson 2017 for a discussion of the kinds of sources selected).

CONCLUSION

As I noted at the beginning of this chapter, data-driven research, and especially transcontextual research, is infinitely more rewarding than anyone who has not done it can imagine. Once we are able to identify which phenomena are the products of good teaching and support and which are national trends that do not significantly vary by institution type, we are in a position to develop pedagogies and fair policies that are likely to achieve our goals. I am hardly the first to bemoan the fact that so many pedagogies, beliefs, and educational policies are based on anecdote or flat-out false beliefs, and I enthusiastically support those who set out to change that situation—including all the contributors to this collection. But as Howard and I have observed in every presentation we have given, and as many authors in this collection also assert, data alone is not sufficient to allow us to understand what is happening when students write from sources. The selection and use of sources is not a simple, linear process, as the new *Framework for Information Literacy for Higher Education* (Association of College and Research Libraries 2015) acknowledges and the Lilac Project is exploring (Starnes and Walker, this volume). We need to know more than simply the frequency with which students do things if we are to understand those actions and respond usefully.

Citation Project research has always had two parts: data analysis and textual analysis (see Howard 2012, 2014; Jamieson 2014; Serviss 2014; Serviss, this volume). The more we have worked with the papers in the Citation Project corpus, the more we have come to realize that this mixed-methods research is essential. The statistical data allows us to plan and execute follow-up research not guided by “something I observed in my class this semester” but by the frequencies and correlations we see in the data. Analysis of papers that move from paraphrase to patchwriting and back, like that shown in Figure 1.3, and those that alternate between quotation and patchwriting with little paraphrase seems to be suggesting there is in fact more than one kind of patchwriting (Jamieson 2014). While the data pointed me to this possibility, it will take close analysis of the papers themselves to discover whether it is true. This research method—using big data to shape close reading—is one we propose for writing studies as the field moves forward, but as I noted at the beginning of this chapter, it is not one that comes easy to all writing studies scholars. I hope that by sharing both the artifacts and the stories of our research, we can help others design—and complete—their own transcontextual projects, moving through the messiness of research design without losing an understanding of the nuance of the text, or what Adam Banks (2015) calls the “messiness of all discourse.”

Note

1. These materials are also available on the Citation Project website (citationproject.net), which also requests that in order to avoid confusion between the initial study and subsequent work, research conducted using these resources should indicate where it replicates Citation Project methods but be called *citation context research*, not Citation Project research (Jamieson and Howard 2013)

APPENDIX 1.A

The Citation Project
Preventing plagiarism, teaching writing

Procedures for Paper coding

PRECODING STEPS

1. Enter your name, today's date, and the paper number in the upper right-hand corner of the Paper Coding Sheet.
2. Enter your name and today's date in the upper right-hand corner of the paper.
3. Note that each source has a number (e.g.: Z0102). Also note that each boxed citation in the paper is numbered in the margin and that number is accompanied by the source number to which the citation refers (e.g.: Z0102). We only code the material inside the boxed areas.
4. On each source, you will find the page number in the upper right-hand corner starting with page 1. Do not use the source's internal pagination in your coding. If the source begins with an abstract generated by the database or with a title page and is otherwise blank, that is numbered as 0 – for example the first page of Z0108.
5. Before you begin coding the paper, review "Glossary for Paper Coding" and "FAQs for Paper Coding."

CODING STEPS

6. To code the paper, read the first boxed area of text and then read the corresponding source and identify the passage to which the citation refers. If the cited material does not appear in the source, do not try to code the material. Simply make a note in the margin and on the coding sheet and move to the next boxed section.
7. Following the definitions on the "Glossary for Paper Coding," highlight all source uses within the box:

pink for summary;
 green for paraphrase;
 yellow for patchwriting;
 blue for copying.

Highlight only cited material inside the boxes.

Note misunderstood or misquoted sources, or other issues of potential concern and or interest, in the margin and in the comment section on the coding sheet.

8. In the second column on the "Paper Coding Sheet," enter the page in the paper where the citation appears. If the citation spans more than one page, enter only the number of the first page (where the boxed area to be coded begins).
9. In the third column of the "Paper Coding Sheet," enter the code of the source used (e.g.: Z0102). The code is written by the boxed area on the paper, and on the first page of the source document itself).
10. In the fourth column of the "Paper Coding Sheet," enter the page number in the source from which the student was working. Use the page numbers written at the top of the PDF, even if this is not the page number cited by the student). If the citation refers to multiple pages in the source or the material spans several pages, enter only the number of the first page used.
11. If there is **only one** type of source use in the boxed area, leave the fifth column of the "Paper Coding Sheet" blank, and write a 1 in one of the last five columns of the "Paper Coding Sheet." (In the case of copying, be sure to check the correct column, depending on whether it was marked as a quotation or not.)
12. If there is **more than one** type of source use in the boxed area, write a 1 in the fifth column of the "Paper Coding Sheet" and determine which type of source use dominates—whichever type has the most words highlighted. Write a 1 in one of the last five columns of the "Paper Coding Sheet."
13. Repeat steps 6-12 for each box of text.

POSTCODING STEPS

14. When you and your coding partner have finished coding the student paper, go through and reconcile. Fill out a new "Paper Coding Sheet" with your agreed-upon codes. In the upper right-hand corner of the new "Paper Coding Sheet," enter both your names, today's date, and the paper number.

APPENDIX 1.B

CITATION PROJECT PAPER CODING SHEET Phase II

Paper Code (e.g.: B01): _____

Coder Name: _____

Coding Date: _____

Fill out these 3 columns			Mark if relevant	Mark an 'X' in one, and only one, of these five columns						
Citation #	Page in the paper	Source used (source #)	Page in source	More than one type of source use in the citation	Copying not marked as quotation	Copying marked as quotation	Patchwriting	Paraphrasing	Summarizing	coder notes (optional)
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										

Sandra Jamieson + Rebecca Moore Howard, 2011

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which maybe viewed at <<http://creativecommons.org/licenses/by-nc-sa/3.0/>>

APPENDIX 1.C

The Citation Project
Preventing plagiarism, teaching writing

Glossary for Paper coding

Accurate synonym

Accurate synonyms are those that do not contradict or change the meaning of the source.

Citation

A signpost embedded in one text and pointing to another. The signpost can come in one or more of the following forms:

1. A signal phrase: *Lincoln charges the living to continue the fight waged by the dead.*
2. A parenthetical citation: *The living must continue the fight waged by the dead (Lincoln).*
3. A footnote or endnote: *The living must continue the fight waged by the dead.*²

Coding

Analyzing and classifying textual features – in this case, how a paper is using its cited source.

Copying not marked as quotation

Reproducing material into a stand-alone sentence, even if it makes minor errors in transcription.

Documentation

Bibliographic information for works that are cited in the paper. This is usually a list of "References" or "Works Cited" at the end of the paper, but in Chicago style the documentation may appear in footnotes.

Ellipses

Used to signal deletions from a passage copied from a source, e.g., "Now we are . . . testing whether that nation or any nation so conceived and so dedicated can long endure."

Keywords

Non-exchangable labels for major concepts. They are "non-exchangable" in that they have no synonyms; they are terms that the source has coined; or they are such standard labels that using a synonym would cause confusion. When used in a source, words and phrases such as *the Middle Ages*, *fundamentalism*, *the White House*, and *deconstruction*, for example, would appropriately be replicated in a paraphrase or summary of that source. What constitutes a keyword is a matter of interpretation rather than rule; hence each coder will have to make these decisions and should make them conservatively: when you are coding and can't decide whether something is a keyword, it isn't.

Paper

The student text that you are coding.

APPENDIX 1.D

The Citation Project Preventing plagiarism, teaching writing

FAQs for source coding

What if the paper attributes something to a source but it isn't in that source?

Don't code it; remove that citation from the numbered list, and renumber the citations that follow it.

What if a paper is paraphrasing sentences from different parts of a source and putting those paraphrases together?

It isn't summary; it's paraphrase.

What if the paper is copying from a source, but with small changes?

- If the changes seem to be typos (transcription errors), code the passage as copying.
- If the changes seem to be deliberate variations, code it as patchwriting.

What if a paper copies subheadings and turns them into a list or sentences?

That's patchwriting, unless quotation marks are used.

How do I code graphics, film, and images?

Because this project is restricted to textual sources, you should not code moving images such as video and film. Do make a note, though, that the paper contains them. Still graphics such as photographs, tables, charts, and graphs should be coded just as any other text. If the paper is generating its own graphics from data elsewhere, you should decide whether the paper is paraphrasing or summarizing. When a paper copies graphics from a source, that is a quotation. Remember to check to see whether the paper cites the source *from which it got the graphics*.

How do I code different uses of a single source *within* a single citation?

Even if more than one type of source use (e.g., copying and paraphrasing) occurs within a single citation, the citation is classified as a single source use. Check the column "More than one type of source use on the citation." Then count the number of words in each source use in the citation and classify the source use according to which source use is greatest (use the coder notes column to make a note of what else is going on in the citation for future study).

What if the paper is misinterpreting its source?

There's no coding category for this, but it would be helpful if you were to note it in the margin of the paper and in the coder notes column on the coding sheet to aid future studies.

What if one source cites another and the student represents it as if the citing source said it?

We're not coding for this, but we would nevertheless like a note about it. Make a note in the margin and the coder notes column, and then code the type of source use.

References

- Anson, Chris M. 2006. "The Intelligent Design of Writing Programs: Reliance on Belief or a Future of Evidence." Plenary address at the annual conference of the Council of Writing Program Administrators, Chattanooga, TN.
- 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.
- Association of College and Research Libraries. 2015. "Framework for Information Literacy for Higher Education." <http://www.ala.org/acrl/standards/ilframework>.
- Banks, Adam. 2015. "Funk, Flight, and Freedom." Chair's address at the Conference on College Composition and Communication, Tampa, FL.
- Bartholomae, David. 1986. "Inventing the University." *Journal of Basic Writing* 5 (1): 4–23. First published in 1985 by Guilford in *When a Writer Can't Write: Studies in Writer's Block and Other Composing Process Problems*.
- Benedicks, Crystal, Kristi Murray Costello, Rebecca Moore Howard, Sandra Jamieson, Kelly Kinney, Tanya Rodrigue, and Tricia Serviss. 2010. "Understanding Students' Use of Sources through Collaborative Research." Half-day workshop presented at the Conference on College Composition and Communication, Louisville, KY.
- Carlson, Jake. 2006. "An Examination of Undergraduate Student Citation Behavior." *Journal of Academic Librarianship* 32 (1): 14–22. <http://dx.doi.org/10.1016/j.acalib.2005.10.001>.
- Creswell, John W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks, CA: SAGE.
- Davis, Philip M. 2002. "The Effect of the Web on Undergraduate Citation Behavior: A 2000 Update." *College & Research Libraries* 63 (1): 53–60. <http://dx.doi.org/10.5860/crl.63.1.53>.
- Howard, Rebecca Moore. 1993. "A Plagiarism Pentimento." *Journal of Teaching Writing* 11 (2): 233–46.
- Howard, Rebecca Moore. 2011. "The Background: Why We Need Data-Driven Research to Understand Plagiarism." Paper presented at the Conference on College Composition and Communication, Atlanta, GA.
- Howard, Rebecca Moore. 2012. "Rhetorical Consequences of Students' Source Choices." Paper presented at the Conference on College Composition and Communication, Saint Louis, MO.
- Howard, Rebecca Moore. 2014. "Why This Humanist Codes." *Research in the Teaching of English* 49 (1): 75–81.
- Howard, Rebecca Moore, and Sandra Jamieson. 1995. *The Bedford Guide to Teaching Writing in the Disciplines*. Boston, MA: Bedford Books.
- Howard, Rebecca Moore, and Sandra Jamieson. 2012. "Take a Deep Breath and Jump: Doing Data-Driven Research When You Aren't Trained in Data-Driven Methods." Plenary address at the Annual Meeting of the Research Network Forum, Convention of the Conference on College Composition and Communication, Saint Louis, MO.
- Howard, Rebecca Moore, Tricia Serviss, and Tanya K. Rodrigue. 2010. "Writing from Sources, Writing from Sentences." *Writing & Pedagogy* 2 (2): 177–92. <http://dx.doi.org/10.1558/wap.v2i2.177>.
- Hull, Glynda, and Mike Rose. 1989. "Rethinking Remediation: Toward a Social-Cognitive Understanding of Problematic Reading and Writing." *Written Communication* 6 (2): 139–54. <http://dx.doi.org/10.1177/0741088389006002001>.
- Hull, Glynda, and Mike Rose. 1990. "'This Wooden Shack Place': The Logic of Unconventional Reading." *College Composition and Communication* 41 (3): 287–98. <http://dx.doi.org/10.2307/357656>.

- Jamieson, Sandra. 2010. "Understanding Students' Use of Sources through Collaborative Research." Paper presented at the Georgia International Conference on Information Literacy, Savannah, GA.
- Jamieson, Sandra. 2012. "Students' Source Choices: A Statistical and Rhetorical Analysis of Researched Papers from Sixteen Institutions." Paper presented at the Conference on College Composition and Communication, Saint Louis, MO.
- Jamieson, Sandra. 2013. "Reading and Engaging Sources: What Students' Use of Sources Reveals about Advanced Reading Skills." Special issue, *Across the Disciplines* 10 (4). <http://wac.colostate.edu/atd/reading/jamieson.cfm>.
- Jamieson, Sandra. 2014. "Understanding Undergraduate and Graduate Students' Source Choices and Uses." Symposia at the Writing Research Across Borders International Conference (International Society for the Advancement of Writing Research), Paris.
- Jamieson, Sandra. 2016. "Is It Plagiarism or Patchwriting? Toward a Nuanced Definition." In *The Handbook of Academic Integrity*, edited by Tracey Bretag, 503–18. Springer. http://dx.doi.org/10.1007/978-981-287-098-8_68.
- Jamieson, Sandra. 2017. "What the Citation Project Tells Us about Information Literacy in College Composition." In *Information Literacy: Research and Collaboration across Disciplines*, edited by Barbara D'Angelo, Sandra Jamieson, Barry Maid, and Janice R. Walker, 117–41. Perspectives in Writing Series. Fort Collins, CO: WAC Clearing House and University Press of Colorado.
- Jamieson, Sandra, and Rebecca Moore Howard. 2013. "Sentence-Mining: Uncovering the Amount of Reading and Reading Comprehension in College Writers' Researched Writing." In *The New Digital Scholar: Exploring and Enriching the Research and Writing Practices of NextGen Students*, edited by Randall McClure and James Purdy, 109–32. Medford, NJ: Information Today.
- Jenkins, Paul O. 2002. "They're Not Just Using Web Sites: A Citation Study of 116 Student Papers." *College & Research Libraries News* 63 (3): 164.
- Kantz, Margaret. 1990. "Helping Students Use Textual Sources Persuasively." *College English* 52 (1): 74–91. <http://dx.doi.org/10.2307/377413>.
- Kennedy, Mary Lynch. 1985. "The Composing Process of College Students Writing from Sources." *Written Communication* 2 (4): 434–56. <http://dx.doi.org/10.1177/0741088385002004006>.
- Lunsford, Karen. 2013. "Replicating Codes: What Does This Mean for Writing Studies?" Paper presented at the Conference on College Composition and Communication, Las Vegas, NV.
- McClure, Randall, and Killian Clink. 2009. "How Do You Know That? An Investigation of Student Research Practices in the Digital Age." *Portal: Libraries and the Academy* 9 (1): 115–32. <http://dx.doi.org/10.1353/pla.0.0033>.
- Nisbett, Richard E., and Timothy D. Wilson. 1977. "The Halo Effect: Evidence for Unconscious Alteration of Judgments." *Journal of Personality and Social Psychology* 35 (4): 250–56. <http://dx.doi.org/10.1037/0022-3514.35.4.250>.
- Pecorari, Diane. 2003. "Good and Original: Plagiarism and Patchwriting in Academic Second-Language Writing." *Journal of Second Language Writing* 12 (4): 317–45. <http://dx.doi.org/10.1016/j.jslw.2003.08.004>.
- Roig, Miguel. 1999. "When College Students' Attempts at Paraphrasing Become Instances of Plagiarism." *Psychological Reports* 84 (3): 973–82. <http://dx.doi.org/10.2466/pr0.1999.84.3.973>.
- Roig, Miguel. 2001. "Plagiarism and Paraphrasing Criteria of College and University Professors." *Ethics & Behavior* 11 (3): 307–23. http://dx.doi.org/10.1207/S15327019EB1103_8.
- Saldaña, Johnny. 2013. *The Coding Manual for Qualitative Researchers*. 2nd ed. Los Angeles, CA: SAGE.

- Serviss, Tricia. 2014. "Understanding Graduate Students' Source Choices and Uses."
Paper presented at the Conference on College Composition and Communication,
Indianapolis, IN.
- Thorndike, Edward L. 1920. "A Constant Error in Psychological Rating." *Journal of
Applied Psychology* 4 (1): 25–29. <http://dx.doi.org/10.1037/h0071663>.
- Titscher, Stefan, Michael Meyer, Ruth Wodak, and Eva Vetter. 2000. *Methods of Text and
Discourse Analysis*. Thousand Oaks, CA: SAGE.