

Andrea A. Lunsford and Karen J. Lunsford

“Mistakes Are a Fact of Life”: A National Comparative Study

This essay reports on a study of first-year student writing. Based on a stratified national sample, the study attempts to replicate research conducted twenty-two years ago and to chart the changes that have taken place in student writing since then. The findings suggest that papers are longer, employ different genres, and contain new error patterns.

Mistakes are a fact of life. It is the response to the error that counts.

—Nikki Giovanni, *Black Feeling, Black Talk, Black Judgment*

Perhaps it is the seemingly endless string of what have come to be called “Bushisms” (“We shouldn’t fear a world that is more interacted”) and the complex response to them from both right and left. Perhaps it is the hype over Instant Messaging lingo cropping up in formal writing and the debate among teachers over how to respond (Farmer 48). Perhaps it is the long series of attempts to loosen the grip of “standard” English on the public imagination, from the 1974 special issue of *College Composition and Communication* (*Students’ Right to Their Own Language*) to a 2006 special issue of *College English* devoted to *Cross-Language Relations in Composition*. Or perhaps it is the number of recent reports, many of them commissioned by the government, that

CCC 59:4 / JUNE 2008

have bemoaned the state of student literacy and focused attention on what they deem significant failures at the college level (see especially the recent reports from the Spellings Commission and Derek Bok's *Our Underachieving Colleges*).

Whatever the reasons, and they are surely complex and multilayered, forms of language use have been much in the news, with charges of what student writers can and cannot (or should and should not) do all around us. The times seemed ripe, then, for taking a close look at a national sample of student writing to see what it might tell us about the current state of affairs. With that goal in mind, we drew up plans to conduct a national study of first-year college student writing and to compare our findings to those of a similar study conducted over twenty years ago.

“The Frequency of Formal Errors,” or Remembering Ma and Pa Kettle

But we are getting a bit ahead of ourselves here. For now, flash back to the mid-1980s. Some readers may remember receiving a letter from Robert Connors and Andrea Lunsford asking them to participate in a national study of student writing by submitting a set of marked student papers from a first-year composition course. That call brought in well over 21,000 papers from 300 teachers around the country, and in fairly short order Andrea and Bob drew a random sample of 3,000 student papers stratified to be representative in terms of region of the country, size of institution, and type of institution.¹ While they later analyzed patterns of teacher response to the essays as well as the particular spelling patterns that emerged (in that study, spelling was the most frequent student mistake by some 300 percent), they turned first to an analysis of which formal errors (other than spelling) were most common in this sample of student writing.

Why the focus on error in the Lunsford and Connors study? Bob and Andrea's historical research had led each of them to caches of student papers with teacher comments focusing on errors that sometimes seemed very out of date if not downright odd (“stringy” syntax, for example, or obsessive comments on how to distinguish between the use of “shall” and “will”), and they wondered what teachers in the 1980s would focus on instead. In addition, the 1938–39 research into student patterns of formal error carried out by John C. Hodges, author of the *Harbrace Handbook of English*, piqued their curiosity—and led to a review of earlier studies. As Connors and Lunsford put it:

Beginning around 1910 . . . teachers and educational researchers began trying to taxonomize errors and chart their frequency. The great heyday of error-frequency seems to have occurred between 1915 and 1935. . . . Our historical research indicates that the last large-scale research into student patterns of formal error was conducted in 1938–39 by John C. Hodges. . . . Hodges collected 20,000 student papers, . . . using his findings to inform the 34-part organization of his *Harbrace Handbook*. (39)

As Connors and Lunsford noted, Hodges did not publish any results of his study in contemporary journals, though in a footnote to the preface of the first edition of his *Handbook*, he did list the top ten errors he found. Connors and Lunsford's research turned up two other "top ten" lists, one by Roy Ivan Johnson in 1917, the other by Paul Witty and Roberta Green in 1930. The three lists are presented in Table 1.²

Table 1: Historical Top Ten Errors Lists

Johnson (1917) 198 papers	Witty & Green (1930) 170 timed papers	Hodges (late 1930s) 20,000 papers
Spelling	Faulty connectives	Comma
Capitalization	Vague pronoun reference	Spelling
Punctuation (mostly comma errors)	Use of "would" for simple past tense forms	Exactness
Careless omission or repetition	Confusion of forms from similarity of sound or meaning	Agreement
Apostrophe errors	Misplaced modifiers	Superfluous commas
Pronoun agreement	Pronoun agreement	Reference of pronouns
Verb tense errors and agreement	Fragments	Apostrophe
Ungrammatical sentence structure (fragments and run-ons)	Unclassified errors	Omission of words
Mistakes in the use of adjectives and adverbs	Dangling modifiers	Wordiness
Mistakes in the use of prepositions and conjunctions	Wrong tense	Good use

Increasingly intrigued to see how formal error patterns might have shifted in the sixty-odd years since these earlier research reports, Connors and Lunsford set out to discover the most common patterns of student errors characteristic

of the mid-1980s and which of those patterns were marked most consistently by teachers. Table 2 presents their findings.³

Table 2: Connors and Lunsford List of Most Frequent Formal Errors

Error or Error Pattern	# found in 3000 papers	% of total errors	# found marked by teacher	% marked by teacher	rank by # of errors marked by teacher
1. No comma after introductory element	3,299	11.5%	995	30%	2
2. Vague pronoun reference	2,809	9.8%	892	32%	4
3. No comma in compound sentence	2,446	8.6%	719	29%	7
4. Wrong word	2,217	7.8%	1,114	50%	1
5. No comma in non-restrictive element	1,864	6.5%	580	31%	10
6. Wrong/missing inflected endings	1,679	5.9%	857	51%	5
7. Wrong or missing preposition	1,580	5.5%	679	43%	8
8. Comma splice	1,565	5.5%	850	54%	6
9. Possessive apostrophe error	1,458	5.1%	906	62%	3
10. Tense shift	1,453	5.1%	484	33%	12
11. Unnecessary shift in person	1,347	4.7%	410	30%	14
12. Sentence fragment	1,217	4.2%	671	55%	9
13. Wrong tense or verb form	952	3.3%	465	49%	13
14. Subject-verb agreement	909	3.2%	534	58%	11
15. Lack of comma in series	781	2.7%	184	24%	19
16. Pronoun agreement error	752	2.6%	365	48%	15
17. Unnecessary comma with restrictive element	693	2.4%	239	34%	17
18. Run-on or fused sentence	681	2.4%	308	45%	16
19. Dangling or misplaced modifier	577	2.0%	167	29%	20
20. Its/it's error	292	1.0%	188	64%	18

As noted above, Table 2 omits spelling errors, which constituted such a large number of the formal errors that Andrea and Bob decided to study them separately (see “Exercising Demonolatry”). In analyzing the other most frequent patterns of formal error and teacher marking of them, Bob and Andrea drew some intriguing conclusions: First, teachers vary widely in their thinking about what constitutes a “markable” error. Second, teachers do not mark as many errors as the popular stereotype might have us believe, perhaps because of the difficulty of explaining the error or because the teacher is focusing on only a few errors at any one time. Finally, they concluded that error patterns had indeed shifted since the time of Hodges’s *Harbrace Handbook*, especially

in terms of a “proliferation of error patterns that seem to suggest declining familiarity with the visual look of a written page” (406).

While Andrea and Bob found errors aplenty in the 3,000 papers from 1984, they also found reason for optimism:

One very telling fact emerging from our research is our realization that college students are not making more formal errors in writing than they used to. The numbers of errors made by students in earlier studies and the numbers we found in the 1980s agree remarkably. (406)

Table 3 presents their comparison of the findings of the three studies.

Table 3: Comparison of Three Studies' Findings

Study	Year	Average Length	Errors per Paper	Errors per 100 Words
Johnson	1917	162 words	3.42	2.11
Witty & Green	1930	231 words	5.18	2.24
Connors & Lunsford	1986	422 words	9.52	2.26

Given the consistency of these numbers, Connors and Lunsford concluded that “although the length of the average paper demanded in freshman composition has been steadily rising, the formal skills of students have not declined precipitously” (406).

Error Studies, 1986–2006

During the two ensuing decades, researchers have continued to study error patterns. Most notable, perhaps, is Gary Sloan's 1990 “Frequency of Errors in Essays by College Freshmen and by Professional Writers,” which found that “[t]he distribution of errors in the students' writing is consistent with figures from previous studies. . . . Connors and Lunsford found 9.52 errors per essay or 2.26 errors per 100 words; my figures for the same are 9.60 and 2.04” (302). Sloan also found that professional writers were prone to making errors, though the errors they made often differed significantly from those of the first-year writers.

During these two decades, researchers also worked to put error study in context. In a 1987 update to Mina Shaughnessy's bibliographical essay on basic writing, Andrea Lunsford reviewed work aimed at reconceiving error as “an active part of learning” (213). And in 1988, Richard Haswell reported on a study of eight error patterns in student writing across the college years and beyond.

Haswell's detailed and carefully nuanced analysis, which studies eight error patterns (formation of possessives, faulty predication, pronoun reference, syntactic parallelism, punctuation of final free modification, end punctuation, punctuation of compound sentences, and orthography) deserves to be read in its entirety. For the purposes of this essay, however, we note that he stresses the need to view all of the categories of error he studies in as rich a context as possible since "the causality of student error is very complex" (495). His own findings suggest that "raw number of errors . . . seems to be growing during college" although, paradoxically, student writers "simultaneously are making measurable growth . . . toward mature competence" (494–95). Thus, Haswell concludes, "To treat surface error as source rather than symptom may still be premature with college age writers" (495). Haswell's findings bear out Lunsford and Connors's analysis of teacher response and marking patterns, which suggests that teachers often mark errors in terms of their relationship to a complex context.

More recently, Christy Desmet and Ron Balthazor are using <emma>, an electronic markup and management application, to conduct what they describe as "a local ethnography of error marking" (6). In a look at errors in "478 essays drawn from ten sections with different instructors," they report "comma errors, development, diction, awkward phrasing, and spelling" as the five most frequent errors.⁴ Studying error in the context of a complex range of factors—or in local contexts, as Desmet and Balthazor are doing—seems a very promising approach to us, and we hope to see more such studies in the coming years. To date, however, such studies are few, and those that have been done repeatedly point to the hurdles researchers faced in bringing them to completion.

A New Study of Student Writing: Those IRB Blues

With this review of error studies as well as the ongoing debate over what constitutes "good" college writing in mind, we set out to replicate the Connors and Lunsford study.⁵ We began the study assuming that the last two decades have ushered in huge changes in writing. To take only the most obvious example, when Bob and Andrea conducted their study, almost all students were writing by hand. Today, students not only use basic word processing but have available many other tools—from color and font type to images and sound—in composing texts. While they write, spell checkers and grammar checkers give them incessant advice. In short, the digital revolution has brought with it opportunities and challenges for writing that students and teachers twenty-two years ago could scarcely imagine. What we had not expected, however,

was the degree to which institutional practices have also changed in twenty-two years. In short, we simply could not have imagined how changes to institutional review board (IRB) policies and procedures would impact our research.

What we quickly learned was that researchers conducting the kind of multi-institutional nationwide research we were attempting must negotiate with the IRB at every single research site involved. Twenty-two years ago, Bob and Andrea's consent forms and other materials easily gained approval from their home institution IRBs—and that approval covered all requirements for institutional review. As a result, their study proceeded apace. For our research, however, while the process of gaining approval from the Stanford and UCSB IRBs was straightforward enough, we found ourselves mired in red tape as we sought volunteers interested in participating in the study; only then could we contact their IRB (if any) asking for permission to proceed. First, we sent out 800 invitations by email, then 2,500 invitations by email, and then 10,000 letters on Stanford University letterhead, and then we appealed to the WPA-L email discussion list—all asking for volunteers to participate and to provide accurate contact information for individual campus IRBs. Upon receiving notices of interest, we began communicating with the local IRB officials. To our surprise, instead of exempting or expediting local approval in light of the Stanford and UCSB approvals, many officials then asked us to go through their own *full review process*. Thus began the tedious, the time-consuming, the mind-numbing task of filling out dozens upon dozens of IRB forms, each with slightly different emphases and questions, and then waiting, sometimes for months, for a response.⁶

Compared to Andrea and Bob's experience, the effects of this expanded and noncollaborative IRB system were chilling. The data collection two decades ago took about three months. For this study, the data collection took six times as long—a full eighteen months. Two decades ago, a much larger number of campuses participated; for this study, we were limited to those that were willing and able to issue approvals in a timely manner.⁷ Two decades ago, each teacher's packet of submitted papers represented nearly the full class;⁸ for this study, the packets typically represented half the number of students. We do not know how many potential teacher participants we lost in this process, although we did hear from some volunteers who found themselves simply overwhelmed when they attempted to sponsor our project on their campuses. Ultimately, our project required literally countless hours of researcher time to submit the same IRB protocol over and over; it unexpectedly required us to trouble several colleagues to become local co-principal investigators; and it

was reviewed and approved of by more IRB committees than ever touch a typical medical study—all so that we can tell you now that in a random stratified sample of 877 (of 1,826 total) anonymous student papers, we found 645 comma splices.

More on Methodology: Preparing the Papers for Coding

But before we could count even one comma splice, before we could pull a random stratified sample, we had to prepare the papers carefully. As the packets arrived, each paper was assigned a unique number that also indicated the region and the type and size of its school in order to yield a stratified sample for analysis (see Appendix 1). As a research assistant marked the sample for stratification, he also removed student names and all other identifying information from all papers.⁹

Although the total paper archive we gathered is—regrettably—smaller than the one in Bob and Andrea’s previous study, it still represents a wide range of papers. Our call for a set of papers from as close to the end of term as possible and one that included teacher comments brought in papers from all types of first-year writing courses (basic, regular, advanced, and specialty courses). And because the data collection lasted eighteen months, we received papers from all college terms, including some summer courses. In the previous study, Lunsford and Connors chose to remove papers that contained formal markers that suggested they were written by students for whom English was a second language, because there were so few of them. For the current study, determining which papers represented ESL writers seemed a harder task, and in any event, we very much wanted to include papers written by multilingual students in our study. The last twenty-two years have seen ESL students more thoroughly integrated into mainstream writing classrooms, and Generation 1.5 students are now recognized as a new group. This decision also helped to broaden the range of papers we could examine. And, finally, many papers came in with their grades on them, and all possible grades from A to F were well represented, again broadening the range of the sample we could study.

Once the papers were rendered anonymous, and once we had a sense of the overall total, we were able to pull our random stratified sample. We used an initial small sample to create the coding rubric and then added to that a large sample to be coded by volunteers. As paper packets continued to arrive and to be processed during the coding phase of the project, we increased the sample size to match the overall stratification.

Developing the Coding Rubric

In attempting to replicate the previous study's development of a coding rubric, we used a procedure employed in the earlier study: Andrea pulled a small, random stratified sample of 25 papers and marked every formal error that she could find. A week or so later, she took another 25 papers and repeated the procedure, marking every formal error. Her results are summarized in Table 4.

Table 4: Formal Errors in a 50-Paper Sample

Error or Error Pattern	# in 50 Papers
Wrong word	79
Comma splice	61
Missing comma after intro word or phrase	55
Possessive apostrophe error	48
Subject-verb agreement error	41
Missing internal citation with page number, etc.	35
Homonym error	32
Missing word	31
Pronoun / antecedent agreement error	28
Fragment	26
Unnecessary comma before coordinating conjunction joining compounds	24
Capitalization (missing or unnecessary)	24
Hyphen (missing or unnecessary)	21
Faulty sentence structure	21
Problems in tense sequence or shift, etc.	20
Pronoun reference error	20
Missing comma or comma needed for restrictive/nonrestrictive elements	17
Spelling error	17
Parallelism error	17
Unnecessary comma before quotation	15
Article error	15
Fused (run-on) sentence	15
Quotation marks inappropriately used for emphasis	13
Colon unnecessary before series or quotation (7) / missing (4)	11
Quotation not introduced or commented on	10
Shift or error in number	9
Wrong or missing preposition	8
Adjective/adverb confusion	8
Shift in person	8

Continued on next page

Table 4: Continued

Missing comma before verbal	7
Missing or extraneous comma in a series	7
Unnecessary comma around prepositional or other phrase	7
Missing quotation marks at beginning or end of quotation or both	7
Quotation not integrated into sentence	7
Dangling or misplaced modifier	7
Missing comma at one or the other end of a phrase	6
Unnecessary comma between verb and object	4
Missing comma with appositive	4
Missing comma before coordinating conjunction joining clauses	4
Missing comma around conjunctive adverbs	3
Unnecessary comma between subject and verb	3
Pronoun case error	3
Idiom error	3
Unnecessary semicolon	3
Split infinitive	2
A/an confusion	2
Comparative adjective error	2

From Andrea's initial list, we pulled the top 25 errors to include in the rubric. One early and clear result of this procedure: it sifted out a group of errors that did not show up at all on Bob and Andrea's earlier top twenty list, an issue we address below.

However, we were also curious about the fate of some of those earlier top twenty errors, and we knew that we needed to compensate for our smaller sample size. As a result, we added to our rubric the errors from the previous study that did not turn up in our initial sample. Additionally, the new rubric asked coders to identify the type of paper, to indicate whether the student or teacher employed technology beyond simple word processing, and to indicate whether the paper was part of a larger portfolio (either multiple drafts of the paper or one element of a whole term's worth of work). If the paper was an item in a portfolio, then the coder was instructed to select the final draft or last assignment to code, as indicated by dates. Again following the methods of the Connors and Lunsford study, our coders tallied both the errors they saw and those that the teacher had marked. To help explain the rubric, we also devised a listing of Error Examples in which we illustrated each error with real student examples pulled from our mini-sample of 50 papers.¹⁰

Coding the Papers and Collating the Results

Now that we had devised and tested a coding rubric of some 40 possible errors, we were ready for the coders. In the previous study, Bob and Andrea trained 50 coders in one afternoon and then, the next day, they worked together to mark 3,000 papers. We estimated that 30 coders would suffice to read 850 papers, and we invited volunteers from among teaching assistants, instructors, and professors in the UCSB programs in writing, education, and ESL/linguistics to help us do so. They arrived at 8:30 a.m. on a chilly Saturday in January 2006. By 10:00, we had all the volunteers trained on using the rubric. As in the original study, although the coders could consult with the researchers about how to classify something they saw in a paper, they were given much autonomy in deciding what constituted an error and how to categorize it. As a result, the judgment calls about specific items were distributed among the group. Armed with stacks of coding rubrics, most coders worked in two large classrooms we had reserved; some retreated to their nearby offices; all read at a diligent pace. Occasionally they returned to the conference room, where fresh papers awaited them along with a smorgasbord of snacks, lunch, and the soothing, continuous burble of two Mr. Coffees and one rescued-from-an-attic Mrs. Tea.

The coders worked feverishly all day, most until 9:30 that night. And it was at that point that we had our first inkling of one of this study's findings: these papers were much longer than those submitted two decades ago. Just how much longer we were to calculate later, after a coder had counted up the pages of body text. That Saturday evening, we quickly reassured our bleary-eyed coders that the 675 papers they had gotten through represented a remarkable achievement.

Then we asked twenty of the volunteers back for Round 2, held two weeks later. In the meantime, more packets of student papers had arrived, so we were able to increase our random stratified sample by 27 papers (for the total 877 papers). Based on feedback from the first reading, we expanded the coding rubric to allow coders to specify more types of comma errors, to list the actual missing and wrong words, and to record semicolon errors separately. Round 2 started at noon on a Friday in early February and went to 8:00 that evening. The remaining 15 papers from that coding session were coded by experienced volunteers over the next workweek. That week, we realized that the tallies for missing commas in the expanded coding rubric suggested that a specific comma error, the missing comma in a compound sentence (MCICS), represented a large proportion of those errors. To determine a better count for those,

we asked a coder to go back through all of the papers again to count just that error.

Once the hurdle of coding all of the papers was behind us, we were faced with the task of adding up all of the error totals. At the end of each coding session, the coders had added their tallies for the set of papers they had marked. (The calculator key chains they used became the study's souvenir.) This gave us a rough estimate of the various totals. We turned then to the task of capturing more accurate totals. Our research assistants gently reminded us that computers were made for more than word processing, and so we decided to enter all of the tally marks into Excel spreadsheets.¹¹ A band of data-entry assistants went to work for three weeks, and our resident expert in Excel compiled all entries into a single spreadsheet. We spent the next few weeks double-checking the entries, then devised a way to extrapolate the totals for the top missing comma error (MCICS), and finally compiled our new top twenty list.¹²

What We Found: Two Major Shifts

Before we turn to a discussion of the particular formal errors in these essays, we want to note two major shifts that have taken place during the last two decades. First, we found that our sense that these papers were quite a bit longer than those in the Connors and Lunsford study was accurate: in fact, these papers turned out to be, on average, over two-and-a-half times longer than those in the previous study. In a further analysis, we found that papers in our sample ranged from a scant 1.5 pages to a densely written 23 pages, and we calculated from the total pages that the average length was 4.15 pages. Assuming the standard 250 words per full page, we calculated that the average number of words was 1,038 per paper. Thus, as Table 5 indicates, research across the decades demonstrates that college student essays have grown longer and longer with time.

Table 5: Comparison of Average Length of Student Essays, 1917–2006

Year	Average Paper Length
1917	162 words
1930	231 words
1986	422 words
2006	1038 words

The second trend we noted is a sea change in the types of papers teachers are asking students to write in first-year writing classes. Although the first study included some reports and a fair number of readings of (mostly) literary texts, the majority of the papers were personal narratives. When we analyzed the kinds of papers represented in this study, we found a range of paper types, as indicated in Table 6.

Table 6: Types of Papers Submitted in 2006

Type of paper	Number found in 877 papers
Researched argument or report	287
Argument with very few or no sources	186
Close reading or analysis	141
Compare/contrast	78
Personal narrative	76
Definition	21
Description	18
Rhetorical analysis	16
Proposal	11
Process analysis	10
Reflective cover letter	3
Other*	30

The "other" category included fiction, letters to aliens, an in-class essay, a news article, several I-searches, a play, several interviews, a biographical sketch, a book report, and several letters.

These results strongly suggest that emphasis on personal narrative has been replaced by an emphasis on argument and research. This finding is supported by Richard Fulkerson's recent map of our discipline, which points to the tremendous growth of argumentation-based textbooks in the last twenty years, despite wide differences in approaches to composition courses (672). Likewise, these results confirm a finding offered by Kathi Yancey and her colleagues: in a national survey of writing programs, an "overwhelming" majority of teachers indicated that they focus on argument- and research-based writing. Together, the two shifts we have identified suggest that student writers today are tackling the kind of issues that require inquiry and investigation as well as reflection and that students are writing more than ever before.

What We Observed about Teacher Comments

We plan to address teacher comments on the papers in our study more thoroughly in a future analysis. At first blush, though, we are struck by how little some things have changed in terms of teacher comments. As they did twenty-two years ago, teachers in this study varied widely in what they decided to mark, and they often focused their marks on just a few specific patterns of error in any one paper. In addition, the vast majority of teachers in this sample (as in the Connors and Lunsford study) marked their papers by hand, employing a variety of inks and pencils and seeming to reserve red ink to signal that students should pay special attention. This time around, many teachers once again chose to mark most often the highly visible and easy-to-circle mistakes, such as apostrophe and spelling errors. Finally, as before, they frequently marked errors that confused a sentence's meaning, such as wrong words.

Generally, the teachers in this study marked slightly fewer of the overall errors than did teachers in the previous study. Two decades ago, the teachers marked 43 percent of the errors the coders found, whereas in this study they marked 38 percent. However, the current teachers were reading papers that were on average two-and-a-half times longer than those in the previous study. In many cases, too, there were references to previous drafts and peer review sessions, and we received 101 papers that were part of portfolios (and in many cases, we received the entire portfolio, too). The comments suggested that many students had already received more extensive feedback on earlier drafts, on which we did not focus in this study. Again, our results support claims by Richard Fulkeron and by Kathi Yancey's team that writing teachers—although they may differ strongly in their theoretical approaches to and aims for first-year composition courses—have nonetheless widely integrated an understanding of writing as a process, along with peer review and reiterative drafting, into their pedagogies.

One surprising finding for us: we received few examples of teachers using specialized computer technologies to comment on student papers. To be sure, we received many examples in which teachers had typed their final comments, and we saw several fairly extensive grading rubrics. What we had expected to collect, though, were examples of teachers using programs such as ConnectWeb, Comment, Daedalus, wikis, blogs, and so on for commenting. Our instructions had explicitly encouraged teachers who use technologies to participate in the study, and we had arranged for technical support. Yet only 56 of the 877 papers had comments that were made via technologies beyond the typed final comment; most typically, they employed Microsoft Word's com-

menting or highlighting features. It is possible that potential participants were put off by the extra steps it would have taken to copy or print the files to submit to us. However, teachers who did submit texts clearly preferred pens and pencils as their commenting technologies. For many teachers, it seems, the electronic commenting tools are still not accessible or convenient enough, or still not pedagogically justified enough, to encourage their use.

What We Also Found: The Most Common Formal Errors in 2006

After nearly two years of data gathering and analysis, we finally have a new list of common formal errors, along with percentages marked by the teachers and by our team of coders, as shown in Table 7.

Table 7: Most Common Formal Errors in 2006

Error or Error Pattern	# found in 877 papers	% of total errors	# found marked by teacher	% found marked by teacher	rank by # of errors marked by teacher
1. Wrong word	3,080	13.7	1,463	48	1
2. Missing comma after an introductory element	2,150	9.6	602	28	5
3. Incomplete or missing documentation	1,586	7.1	722	46	3
4. Vague pronoun reference	1,495	6.7	405	27	8
5. Spelling error (including homonyms)	1,450	6.5	788	54	2
6. Mechanical error with a quotation	1,444	6.4	674	47	4
7. Unnecessary comma	1,175	5.2	346	29	10
8. Unnecessary or missing capitalization	1,168	5.2	490	42	7
9. Missing word	1,024	4.6	491	48	6
10. Faulty sentence structure	996	4.4	297	30	13
11. Missing comma with a nonrestrictive element	850	3.8	229	27	16
12. Unnecessary shift in verb tense	847	3.8	304	36	11
13. Missing comma in a compound sentence	814	3.6	300	37	12
14. Unnecessary or missing apostrophe (including <i>its/it's</i>)	693	3.1	372	54	9
15. Fused (run-on) sentence	668	3.0	189	28	18
16. Comma splice	645	2.9	257	40	14
17. Lack of pronoun-antecedent agreement	612	2.7	253	41	15
18. Poorly integrated quotation	612	2.7	154	25	19
19. Unnecessary or missing hyphen	562	2.5	152	27	20
20. Sentence fragment	531	2.4	223	42	17

Two items on this list instantly leapt out at us. First, while spelling errors outnumbered all others in the Connors and Lunsford study by three to one, spelling errors in our study came in at number 5, accounting for only 6.5 percent of all errors found. Second, “wrong word” in our study was by far the most frequent formal error, accounting for nearly 14 percent of the errors. These findings are dramatic and, it seems to us, related in interesting ways. Since almost every one of our 877 papers was word processed (a very few were handwritten on loose-leaf paper), we assume that the spell-check function took care of many potential spelling problems. Indeed, a great number of the spelling errors in our study are homonyms and proper nouns, mistakes that spell-checkers understandably do not flag.

But every blessing brings its own curse. In this case, many of the wrong word errors appear to be the result of spell-checker suggestions. A student trying to spell “frantic,” for example, apparently accepted the spell-checker’s suggestion of “fanatic.” Wrong word for sure. In addition, some students appear to be using a thesaurus feature without also using a dictionary to understand the nuances of meaning for various words—“artistic,” for example, when “aesthetic” is the appropriate choice. Still other wrong word mistakes seem to result from choosing a word that has a somewhat similar sound: “concur” rather than “conclude” or “analyses” rather than “analyzes.” Finally, many wrong words seem to come from the simple failure to proofread: writing “begging” for “beginning” is no doubt such a case in point.

A second category of mistakes also surfaced early on. Andrea’s preliminary analysis of 50 papers had turned up a number of problems with sources and attributions. Indeed, these errors came to the fore in our coding, ranking numbers 3, 6, and 18 in our list of top twenty formal errors.¹³ The shift to research-based and argumentative writing clearly accounts for many of these mistakes: as we read the 877 papers, we noted students struggling with the use of sources on every front, from omitting citations completely to documenting them in idiosyncratic if not downright bizarre ways. Such struggles seem to us a natural and necessary part of the practice that students must do to become familiar with, much less master, any one documentation style: after all, entering the conversation in a field, showing that you know the issues and have something to contribute to them, choosing among a huge range of possible sources, and using them to document the work related to any particular topic are not easy skills to develop, especially for novice writers. In any case, teachers spent a lot of energy on correcting such errors, marking half of all missing or incomplete documentation mistakes, for example. It stands to reason that

instructors would be attentive to such problems in research-based essays and arguments: after all, using sources appropriately and citing them clearly are major parts of such an assignment.¹⁴

Other mistakes—especially the number of capitalization and hyphenation errors—initially puzzled us, though a little reflection suggested a number of possible explanations. Some capitalization errors, for example, appear to result from Word automatically capitalizing a word that follows a period (such as a period used with an abbreviation). In these cases, the student had not corrected the error, even though it could have been caught with careful proofreading. In still other cases, students seemed inclined to use capitals for the subject terms of their research papers, words that for them seemed to take on a certain significance. For instance, a student capitalized every occurrence of “basketball,” as well as the names of other sports, in a paper that began, “Basketball, America’s other pastime, is the third most popular sport in the USA after Baseball and Football.” Alternatively, in other cases students seemed to be treating objects as proper nouns. For example, in this excerpt, the student visually equates the GED with a high school diploma through capitals: “One common belief is that a person with a GED education is less educated than a person who has achieved a High School Diploma”—implying that a diploma would be an “HSD.” Surely the internal caps in names like eBay or iPod may add to student confusion—such seemingly random capitalization cannot help students who are trying to figure out why some words are capitalized in formal writing and others are not. Yet the most prominent uses of initial caps are advertising and headlines, and perhaps these kinds of texts may be contributing to the increase in capitalization errors. For many students, headlines and slogans may be very common reading—a big part of the nonrequired reading they do.

Hyphens also seemed to be causing a good bit of confusion. A number of hyphen mistakes apparently come from students hyphenating two-part verbs—“put-up,” “log-in,” “shut-down.” Students seem more puzzled than ever by the fact that “sign-up sheet” is hyphenated but “sign up here” is not. In general, however, the trend seems to be moving more toward one unhyphenated word—firewall, laptop, email. To be sure, conventions regarding “email” are still very much in flux, with some reference works stipulating the hyphen and others, increasingly, rejecting it. This trend toward loss of hyphenation also seems apparent in formerly hyphenated prefixes (“supermarket,” “overeat”) and with compound adjectives. Our dictionaries list hundreds of “anti”-compounds, and the only hyphenated ones involve proper nouns (“anti-Russian”) and words

that start with “i” (“anti-immigration”). Could we have a “great hyphenation migration” on our hands? No wonder students are confused.

Of all the errors we noted, those we termed “faulty sentence structure” intrigued us most. Some of these errors seem to arise when students cut and paste passages from one sentence to another, or when they draft a sentence and then delete part of it to correct a mistake—but do not delete enough. But we found many more “faulty sentence structure” errors than these reasons could account for, so much so that we speculate that a number of them may result from students attempting to address complex topics in complex ways. Perhaps the rise in the number of these errors signals the cognitive difficulty associated with argument- and research-based writing, as might be expected to accompany a shift from personal narrative to argument and research.

In any case, faulty sentence structures certainly caused our readers to pause and say “What?!?” These errors tended to attract both teachers’ and coders’ attention because they so often confounded meaning. Critics of current educational practices point to sentences such as these as signs that today’s digital texts are undermining clear thought and that high schools are failing to sharpen syntactic skills. But let’s think again. Consider, for instance, the sentence that served as one of the examples of faulty sentence structure for the coders in our study: “However, Marlow had put caps in the gun, proving that Carmen became infuriated because she was rejected by Regan, as Marlow had also done, and killed Rusty.” Faulty sentence structure, yes. But it is worth reflecting on what may lie behind errors such as these, which may actually signal syntactic growth. This sentence, for instance, is attempting to do some hard work: to signal the temporal and causal relationships among different scenes from *The Big Sleep*. It comes from a thoughtful analysis of that novel,¹⁵ and even though this sentence is overwhelmed by the many incidents that occur, it is trying to sum up a very complex narrative. “It’s the response to error that counts,” as Nikki Giovanni reminds us: when we find examples of such fractured syntax, then, it seems especially important to respond in an open and exploratory way, searching with the student writer for the intended meaning.

How Our Findings Differ from the Connors and Lunsford Study

Seven student errors from Andrea and Bob’s study dropped off in ours (though they do appear farther down in our list, as runners-up to the top twenty): wrong or missing verb ending; wrong or missing preposition; unnecessary shift in pronoun; wrong tense or verb form; lack of subject-verb agreement; missing

comma in a series; and dangling modifier.¹⁶ Somewhat surprising to us were the small number of dangling modifiers we found, and while it is conceivable that a few of these were counted as “faulty sentence structure,” we believe it will take a more fine-grained analysis than the one we have done here to discover what changes in sentence length and syntax might be related to this admittedly small shift—dangling modifiers ranked number 19 in the Connors and Lunsford study. A few of the other errors that dropped out—wrong verb forms, missing verb endings, and subject-verb agreement—are sometimes flagged by grammar checkers, but again we doubt that these checkers alone can be responsible for the reduction in the number of such errors. When we looked at the wrong or missing prepositions in our study, we found that a number of these were counted as “wrong words,” which would account for this change. Finally, the missing comma error suggests to us that students are still struggling with commas, though they are doing so with slightly different patterns than turned up twenty-two years ago.

What We Did Not Find

At the opening of this essay, we noted the many hard-core worriers who see a precipitous decline in student writing ability and who often relate that decline to the creeping influence of IM and other digital lingo as well as to sliding standards. Our findings do not support such fears. In fact, we found almost no instances of IM terms (“gtg,” “imho”) or even smilies in students’ formal prose, although they sometimes appeared in notes to teachers or in the peers’ comments. The students in this sample seemed aware of the ancient principle of *kairos* and wrote with a sense of what is appropriate for formal college writing. More surprising was the little evidence of what has come to be called—perhaps in homage to Winston Weathers’s charming and important *An Alternate Style*—alternate or alternative discourse. With the exception of a handful of funny and often imaginative letters to aliens, all from the same class, as well as some fiction, the papers we examined stuck resolutely to what Weathers dubbed Grammar A: traditional usage, organization, and style. We had imagined, given our field’s lively and intense discussion of alternate styles in the last decade, that we would see more evidence of such experimentation in student writing today.

We also found very little use of the many tools available to student writers today. To be sure, two of the essays were dressed to the nines with superimposed images, clip art, and wildly colored fonts—and a few included tables, charts, or figures. But only 25 papers used images at all, and only 5 more used

colored fonts, hyperlinks, or blog-style entries. For the most part (847 of 877, to be exact), these student writers were not illustrating their texts, nor were they making use of different type sizes, fonts, color, and so on, much less making use of sound or video. No Web texts were submitted to our study. This finding may suggest that teachers' assignments do not yet encourage the use of such tools, or that teachers and students do not have ready access to the technologies that would support their use in writing. We suspect, however, that student writers simply do not yet associate such tools with formal school writing. In any event, for all the attention we give to multimedia forms of writing in our own teaching, and for all the advances the field of writing and rhetoric has made in teaching writing with technology, student writers in these first-year college classes continue to produce traditional print-based texts.

One More Word on Error

As we noted earlier, studies of error across the last ninety years yield remarkably similar findings. This conclusion holds true for our study as well, as shown in Table 8.

Table 8: Comparison of Error Rates per 100 Words

Study	Year	Errors per 100 Words
Johnson	1917	2.11
Witty & Green	1930	2.24
Connors & Lunsford	1986	2.26
Lunsford & Lunsford	2006	2.45

In comparing these numbers, we note that the Connors and Lunsford study did not include spelling errors, since they were the subject of another study. Our study did include spelling errors: if those are excluded, then the rate of error per 100 words (i.e., 2.299) remains almost exactly the same as it has been during the last century, though types of error vary considerably.

In looking at the rate of error in our and other studies, we are reminded of Joseph Williams's essay "The Phenomenology of Error." In its published form, that article contained 100 deliberate formal errors. Most readers, however, were not aware of them until the final sentence—which dramatically announced their presence. Noticing errors depends, then, on the reader's context. As Williams argues, if the piece of writing is professional prose, and if it is cognitively challenging and interesting, then readers do not notice error. The rate of error

in our study, then, should also be seen as rate of *attention* to error. When readers look for errors, they will find them. For the current study, our coders were looking for 40 different types of errors, and they found an awful lot of them. Even so, the rate of error in this study remains consistent with results across nearly 100 years. Those who believe that we ought to be able to eliminate errors from student writing may need to realize that "mistakes are a fact of life" and, we would add, a necessary accompaniment to learning and to improving writing.

Conclusion

We offer these findings with caution, for the total number of papers we received—despite the extraordinary effort and generosity of our volunteers—was much smaller than we had hoped. Yet the study does reveal several important trends, such as the dramatic increase in length of student writing and the shift in the kinds of assignments instructors are giving, assignments that lead to concomitant shifts in errors. Perhaps most important, contrary to what the doomsayers would have us believe, this study confirms that the rate of student error is not increasing precipitously but, in fact, has stayed stable for nearly one hundred years. Nor, really, does it make sense to expect that today's students ought to make fewer mistakes as they learn to write than did their predecessors. The last two decades have seen massive changes in student enrollments, revolutions in writing technologies, and a nationwide shift in first-year writing courses to genres that demand particular cognitive and rhetorical strategies. In the face of these changes, student errors are not more prevalent—they are only *different*.

Our task now is to understand and document those differences better—to continue to work toward a more nuanced and context-based definition of error; to see whether similar large-scale studies produce similar top twenty lists; to identify any significant differences across regions, various groups, or disciplines; to ask how we might adapt our technologies to reduce certain errors and how we might adapt our pedagogies to address the errors to which technologies contribute; to analyze whether our focus on academic discourse is paying off in WAC issues and to compare writing teacher markings with those of teachers in other disciplines. One study cannot provide the documentation needed to convince administrators of the worth of college writing courses, or to demonstrate to colleagues the need to look beyond their own anecdotal accounts of student error, or to make visible the very interesting

shifts that occur in each generation of college writers. Rather, we need a coordinated agenda.

Whether we can coordinate efforts in ways that allow us to meet IRB requirements seems a huge question arising from this study. Instead of two researchers attempting to conduct a centralized nationwide study, we might turn to a process used in many scientific fields. In this process, local researchers conduct experiments and then deposit the results into a central database, which in turn aggregates and creates models or visual representations of the data. We imagine a nationwide effort in which local writing program administrators (WPAs) could be principal investigators (PIs) for their own campuses, conduct a local version of this study, and then submit the anonymous results to a central location (perhaps sponsored by CCCC?). Or perhaps a consortium of WPAs could conduct a comparative study regionally, the better to make the data anonymous. Perhaps those campuses that already have large archives of student papers—often digital archives—could lead the way. In addition, if PIs could gain permission, as we did, to deposit collected student papers into a national archive (e.g., University of New Hampshire's), our field could build an important historical record.

Whether our field can move to adopt any of these suggestions or not, this study reaffirms our belief that student writers and the work they produce are worth such efforts—and more. As a group, the 877 papers we read were attempting to address serious issues in serious, if still maturing, ways; they radiated good humor and found amusement in things large and small; they wrestled with difficult sources and with textual conventions of all kinds; and they documented a range of contemporary values as well as hopes and dreams. And yes, they made mistakes—some real whoppers, others only tiny missteps, but all of them asking for our careful response.

Appendix 1: Stratification of All Student Papers

Region	1	2	3	4	5	6	7	Total
Total number of papers	316	292	317	296	158	118	329	1826
Total number of packets	24	20	18	18	12	8	19	119
Total number of 4-year schools*	16 (238)	15 (217)	13 (249)	9 (126)	7 (106)	5 (52)	10 (151)	75 (1139)
Total number of 2-year schools	8 (78)	5 (75)	5 (68)	9 (170)	5 (52)	3 (66)	9 (178)	44 (687)
Total number of state schools	17 (195)	11 (153)	13 (231)	12 (210)	8 (84)	6 (90)	15 (275)	82 (1238)
Total number of private schools	7 (121)	9 (139)	4 (68)	6 (86)	4 (74)	2 (28)	4 (54)	36 (570)
Total number of proprietary schools	0 (0)	0 (0)	1 (18)	0 (0)	0 (0)	0 (0)	0 (0)	1 (18)
Number of schools with total enrollments under 1,000	0 (0)	2 (30)	1 (1)	2 (37)	0 (0)	0 (0)	2 (27)	7 (95)
Enrollment 1,000–3,000	4 (46)	5 (86)	1 (15)	4 (42)	3 (37)	3 (37)	2 (26)	22 (289)
Enrollment 3,000–5,000	6 (74)	7 (82)	3 (60)	1 (16)	3 (62)	3 (27)	3 (69)	26 (390)
Enrollment 5,000–10,000	6 (102)	4 (59)	6 (103)	3 (54)	2 (20)	1 (46)	4 (64)	26 (448)
Enrollment 10,000–20,000	4 (42)	1 (24)	4 (59)	3 (44)	3 (26)	1 (8)	3 (50)	19 (253)
Enrollment over 20,000	4 (52)	1 (11)	3 (79)	5 (103)	1 (13)	0 (0)	5 (93)	19 (351)

*Note that the first number refers to the number of packets of papers received; the number in parentheses indicates the number of papers.

Notes

1. Connors and Lunsford published several essays resulting from this study, including “Frequency of Formal Errors in Current College Writing, or Ma and Pa Kettle Do Research.” This essay, which attempted a bit of humor—a try at what is now called “alternative discourse”—was huge fun to write. The loss of Bob Connors in a motorcycle accident in the summer of 2000 has left the field of what he termed composition-rhetoric without one of its best (and often funniest) spokespersons. We dedicate this essay to his memory.
2. The lists are quoted in Connors and Lunsford (405). The terms have been updated where possible.
3. The sample of papers Connors and Lunsford analyzed appeared to be written by native speakers. As a result, they pulled out of the sample any essays that were characterized by obvious ESL markers—though there were very few of these.
4. Desmet and Balthazor found the following errors: Comma errors, Development, Diction, Awkward, Spelling, Documentation, Other punctuation, Wordy, Apostrophe, Paragraph coherence, Tense, Transition, Comma splice, Expletive construction, Agreement pronoun-antecedent, Coherence, Paragraph unity, Vague pronoun reference, Agreement subject-verb, Fragment, Passive voice, Wrong preposition, Organization, and Logical fallacy—in decreasing order of frequency.
5. We are grateful to Bedford/St. Martin’s for sponsoring both the previous and the current study. In particular, we thank editors Carolyn Lengel and Stephanie Butler for their patience, hard work, and pragmatic advice throughout this project.
6. Many thanks to James Ford for his help in this phase of the research.
7. Throughout the data collection, too, we were reminded several times that a natural disaster is a national disaster, with rippling effects on our research that we did not anticipate. When hurricanes Ivan (2004), Katrina (2005), and Rita (2005) destroyed communities, they put an enormous strain on nearby academic institutions that welcomed people who were displaced. With regret, some colleagues were forced to put aside their participation in our study, either because their own schools were flooded or because their programs suddenly doubled in size. Some of the student papers we did receive likewise attest to the trauma that these experiences evoked.
8. For the earlier study, teachers also sent multiple classes’ worth of graded papers. In the current study, sometimes teachers from the same school sent their papers in the same packet.
9. We had collected papers with names on them so that we could allow students and teachers a window of time in which to withdraw from the study, should they wish; this procedure was a safeguard against coercion. The most common type of “other identifying information” removed was the phone number, especially as stu-

dents informed teachers and peer group members how to reach them.

10. Many thanks to Alison Bright, who helped identify these examples and, at this point, became the project coordinator.

11. Many thanks to Elizabeth Freudenthal, who patiently led the data-entry and fact-checking, and to Paul Rogers, who contributed his Excel expertise.

12. The *missing comma in a compound sentence* errors (MCICS) came to be fondly known as the McIcks by our research team. Here is how the extrapolation was devised: We knew from the Round 2 rubric what proportions of all missing comma errors the MCICS constituted for both coders and teachers. We applied those percentages to the total missing comma errors on the Round 1 rubric, and we extrapolated totals for both coders and teachers for the Round 1 MCICS. Then we added the Round 1 extrapolation to the count from Round 2.

We decided to use this extrapolation rather than the actual count that the single coder found because, when the coder focused on just one error instead of 40 possible errors, she found an even higher percentage of MCICS than our regular coders or the teachers did.

13. Note, too, that we did not count the errors on the Works Cited pages; had we done so, the number would have been much, much higher.

14. Of the papers that cited resources, 440 used MLA style; 42, APA style; and 29, another style that, if it was recognizable, was usually CSE.

15. Although we might guess that the student had also seen the film, the paper analyzes the novel.

16. In addition, we note that *its/it's* would have dropped off the list if we hadn't combined it with unnecessary or missing apostrophes. It was a separate error in the Connors and Lunsford study, coming in at number 20.

Works Cited

- Bok, Derek. *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More*. Princeton, NJ: Princeton UP, 2005.
- Bush, George W. Presidential address. Washington, DC, June 27, 2006.
- Connors, Robert, and Andrea A. Lunsford. "Frequency of Formal Error in Current College Writing, or Ma and Pa Kettle Do Research." *College Composition and Communication*, 39.4 (1988): 395-409.
- Desmet, Christy, and Ron Balthazor. "Finding Patterns in Textual Corpora: Data Mining, Research, and Assessment in First-year Composition." Paper presented at Computers and Writing 2006, Lubbock, Texas, May 25-29, 2006.
- Farmer, Robert. "IM Online. RU?" *Educause Review*, 40.6 (2005): 48-63.
- Fulkerson, Richard. "Composition at the
- Cross-Language Relations in Composition*. Spec. issue of *College English* 68.6 (2006).

- Turn of the Twenty-First Century.” *College Composition and Communication*, 56.4 (2005): 654–87.
- Giovanni, Nikki. *Black Feeling, Black Talk, Black Judgment*. New York: W. Morrow, 1970.
- Haswell, Richard. “Error and Change in College Student Writing.” *Written Communication* 5 (1988): 470–99.
- Hodges, John C. *Harbrace Handbook of English*. New York: Harcourt, 1941.
- Johnson, Roy Ivan. “The Persistency of Error in English Composition.” *School Review* 25 (October 1917): 555–80.
- Lunsford, Andrea. “Basic Writing Update.” *Teaching Composition: Twelve Bibliographical Essays*. Ed. Gary Tate. Texas Christian UP, 1987: 207–27.
- Lunsford, Andrea, and Robert J. Connors. “Exercising Demonolatry: Spelling Patterns and Pedagogies in College Writing.” *Written Communication* 9 (1992): 404–28.
- Sloan, Gary. “Frequency of Errors in Essays by College Freshmen and by Professional Writers.” *College Composition and Communication* 41.3 (1990): 299–308.
- Students’ Right to Their Own Language*. Spec. issue of *College Composition and Communication* 25 (1974).
- United States. Dept. of Education. Spellings Commission. *Commission Report: A National Dialogue: The Secretary of Education’s Commission on the Future of Higher Education*. August 9, 2006 draft. 13 Sept. 2006. <<http://www.ed.gov/about/bdscomm/list/hiedfuture/reports.html>>.
- Weathers, Winston. *An Alternate Style: Options in Composition*. Rochelle Park, NJ: Hayden, 1980.
- Williams, Joseph. “The Phenomenology of Error.” *College Composition and Communication* 32.2 (1981): 152–68.
- Witty, Paul A., and Roberta La Brant Green. “Composition Errors of College Students.” *English Journal* 19 (May 1930): 388–93.
- Yancey, Kathleen Blake, Teddi Fishman, Morgan Gresham, Michael Neal, and Summer Smith Taylor. “Portraits of Composition: How Composition Gets Taught in the Twenty-first Century.” (forthcoming)

Andrea Abernethy Lunsford

Andrea Abernethy Lunsford is the Louise Hewlett Nixon Professor of English and director of the Program in Writing and Rhetoric at Stanford University. She writes collaboratively whenever possible and is now at work on two large collaborative projects: *The Sage Handbook of Rhetoric* and *The Norton Anthology of Rhetoric and Writing*.

Karen J. Lunsford

Karen J. Lunsford is an assistant professor of writing in the Writing Program at the University of California, Santa Barbara. She studies how interactions among people, institutions, media, and practices define what counts as acceptable argumentation for a community. She is currently studying distributed publication systems in the sciences.