Analysis of Student Revisions on a State Writing Test

Lindy Crawford, Susan Lloyd and Kelly Knoth

Assessment for Effective Intervention 2008; 33; 108
DOI: 10.1177/1534508407311403

Assessment for Effective Intervention
Volume 33 Number 2
March 2008 108-119

10.1177/1534508407311403
http://aei.sagepub.com
hosted at
http://online.sagepub.com

Analysis of Student Revisions on a State Writing Test

Lindy Crawford
University of Colorado at Colorado Springs
Susan Lloyd
Fairfax County, Virginia
Kelly Knoth
University of Colorado at Colorado Springs

Type and quality of revisions made by students between first and final drafts of a state writing test were scored using a revision taxonomy. Scorers categorized revisions first by unit (e.g., word, phrase, sentence), and then by type (e.g., addition, substitution, spelling). They then evaluated the impact of each revision on the readability of the final draft as "improved," "decreased," or "neutral." Descriptive analyses revealed three major findings. First, in terms of unit changes, word revisions were the most prevalent (approximately 40%) across students in Grades 5 and 8. Second, when coding for type of revision, substitutions accounted for approximately 45% of all revisions, while revisions classified as "additions" had the most positive impact on students' final drafts (i.e., positively impacting readability 66% to 76% of the time). An interesting third finding was the relationship between rate of writing and score on the writing test: Students in special education with Individualized Education Programs goals in writing made approximately the same number of revisions per 100 words as students in the general population but wrote 100 words less, on average.

Keywords: *student revisions; state achievement tests; writing fluency; writing assessment*

Most state tests include a direct assessment of students' writing. Students are provided one or more prompts, and their task is to construct a response to the prompt(s) within a given amount of time. Differences exist across various states related to how much weight this portion of the test is given when calculating a total score. For example, in Colorado, constructed response items account for 36% of a student's total writing score (students complete one extended response and three short constructed responses; Colorado Department of Education, 2005), whereas in the Delaware Student Test Program, students complete two extended writing prompts (one text-based and one stand-alone), comprising 100% of their total score on the writing test (Delaware Department of Education, 2006). This increased emphasis on direct assessment of writing on large-scale tests has been significantly influenced by three decades of research and practice supporting the process model of writing as first articulated by Flower and Hayes (1981).

An emphasis on stages of writing (i.e., planning, writing, revising, and rewriting) within the process model has helped shape the format of many state assessments, so that students are now provided with two or more sessions in which to engage in these different stages. Important for this study is the relatively untested assumption that providing students with time to revise their first drafts will result in improved final compositions. This assumption is based on what we have learned about students' writing during classroom instruction but has little empirical support in the context of state assessments. In this study we investigate this assumption by analyzing the breadth and depth of student revisions made during a state assessment of writing to answer three related questions: (1) What kind of revisions do students make on state assessments? (2) How do revision behaviors differ across grade level (fifth and eighth grade) and educational classification (students with and without Individualized Education Programs [IEP])? And (3) Is there a relationship between

Authors' Note: This research was funded through a postdoctoral fellowship sponsored by the National Academy of Education and the Spencer Foundation.

Address correspondence to Lindy Crawford, PhD, University of Colorado at Colorado Springs, College of Education—COH6, 1420 Austin Bluffs Parkway, P.O. Box 7150, Colorado Springs,CO 80907-7150; e-mail: mcrawfor@uccs.edu.

the quantity of revisions made on a first draft and the score earned on a final draft of a state assessment?

Review of Research

Revision as a Process

The early work of Rohman and Wlecke (1964) described writing as unfolding in linear stages, beginning with prewriting and ending with a final draft. In the 1970s, seminal work by Flower and Hayes (1981) resulted in a much richer framework for interpreting writing. Termed the *cognitive process model of writing*, this conceptualization provided a theoretical framework through which the processes of writing could be understood and analyzed. Importantly, in the Flower and Hayes model, revising was described as much more than mere editing and was viewed as a fluid and recursive process that may occur between successive drafts but may also occur as a writer first puts thoughts into words. In its truest sense, revising involves more than a mere surface-level edit of words and phrases, instead representing a thoughtful analysis of what is written and an evaluation of the success of the writing in conveying its intended message. It has been described as a decision-making process that "depends on a dynamic interplay of knowledge and intentions" (Flower, Hayes, Carey, Shriver, & Stratman, 1986, p. 20).

To successfully revise a piece of writing, writers must engage in two different subprocesses. First, they must be able to identify the problems or weaknesses in their rough drafts, and second, they must be able to correct these perceived problems in their final drafts (Beal, 1987; Scardamalia & Bereiter, 1983). According to Fitzgerald (1987), this is "a cognitive problem-solving process in that it involves detection of mismatches between intended and instantiated texts, decisions about how to make desired changes, and making the desired changes" (p. 484). Some evidence suggests that young or inexperienced writers struggle in the first stage, the ability to recognize a problem, and not in their ability to fix the problem once it has been detected (Beal, 1990). Other researchers have found that students with disabilities face challenges both with detecting errors and with correcting those errors on detection (Graham, 1997).

Quantity and Quality of Revisions

A substantial amount of research into the revision behaviors of children, adolescents, and college students has been conducted over the past 30 years. General conclusions can be drawn from a review of this research, including the following: (a) differences exist between novice and expert writers in terms of breadth, depth, and quality of their revisions (Fitzgerald, 1987); (b) good writers revise as they write and not necessarily during artificially imposed breaks (Bridwell, 1980; Stoddard & MacArthur, 1993); and (c) when provided with an explicit strategy for revising, final drafts of students with and without disabilities show significant improvement (Daiute & Kruidenier, 1985; De La Paz & Graham, 1997; Wong, Butler, Ficzere, & Kuperis, 1996).

Revision analyses pertaining specifically to the writing of younger students have yielded fairly consistent results. For example, studies have shown that beginning writers do not engage in "spontaneous revising" (Berninger & Swanson, as cited in Berninger, 2000, p. 72), nor do their revisions succeed in improving text at multiple levels of language (Whitaker, Berninger, Johnston, & Swanson, 1994). Furthermore, the process of revision for younger or inexperienced writers does not tend to significantly improve the quality of their final products (Fitzgerald, 1987; Matsumura, Patthey-Chavez, Valdés, & Garnier, 2002), nor does the process often extend beyond surface-level changes or basic editing (Perez, 2000).

Similar findings as those discussed have been reported about the revision behaviors of students in special education. First, for writers receiving special education services, revisions do not always improve the quality of the final text (MacArthur & Graham, 1987) and in one study increased the quality of final compositions less than 50% of the time (MacArthur, Graham, & Schwartz, 1991). Furthermore, revisions made by students in special education are primarily mechanical in nature (e.g., handwriting, punctuation, recopying of text) and do not impact overall meaning of the text (Graham, 1997; Graham, MacArthur, & Schwartz, 1995).

Revisions on State Assessments

Revision behaviors of students on large-scale assessments have not been studied as extensively as revisions made by students in instructional settings or during highly controlled intervention studies. Some evidence exists, however, that highlights students' tendencies to make surface-level changes in test settings (National Assessment of Educational Progress, 1986). Goldberg, Roswell, and Michaels (1995–1996) reported that the revision behaviors of students in Grades 3, 5, and 8 on the Maryland School Performance Assessment Program were relatively minor and more surface level than substantive. Goldberg et al.'s contextual analysis of student revisions (made in response to a peer review process built into the testing situation) revealed a small but positive relationship with scores obtained on a language use measure (mechanics, grammar, etc.) but little impact on scores obtained on a 4-point holistic scale of writing quality. They also found that the revision behaviors of

Table 1
Student Demographics

Demographic	Grade 5	Grade 8
Gender		
Male	97	95
Female	70	75
Unknown	0	1
Ethnicity		
American Indian	5	0
Asian	11	7
African American	4	10
Hispanic	10	15
White	137	138
Unknown		1
Special education classification		
Learning disabled	20	21
Mentally retarded	4	8

students differed across grade levels. For example, the most prevalent change at third grade was converting a printed rough draft into a cursive final draft. Other changes included word and phrase substitutions that generally maintained the meaning of the text. Students in Grade 5 also made many surface-level revisions but tended to make more elaborate revisions, including additional detail that extended the message contained in the rough draft. Interestingly, at Grade 8, revisions were most often minor and did not impact the writer's overall message.

As two prominent reviews of the research make clear, we have a solid understanding of the kinds of revisions students make (Fitzgerald, 1987) and the effects of instructional interventions on the quantity and quality of these revisions (Gersten & Baker, 2001), but the context for the majority of this research has been the classroom. Furthermore, the number of students participating in most of these studies has been limited, possibly due to the time-consuming nature of content analysis and the demands of intervention research. We do not know if these findings generalize to large-scale state assessments, nor if they generalize across a larger group of students. In this study, therefore, we conducted an in-depth content analysis of the revision behaviors of 342 students in general and special education across Grades 5 and 8 on a state assessment of writing.

Methodology

Sample

Eight elementary and three middle schools were randomly selected across one relatively large urban school district (N = 19,894) in the western United States. The sampling plan allowed for inclusion of approximately 35 students at each school site. A larger sample was not included (nor needed) due to the complexity of the anticipated analyses. State test writing samples completed by these students were selected using a stratified random sampling plan based on students' previous year's proficiency scores on the state assessment. We oversampled in special education by randomly selecting 20 additional tests to generate the sample size needed for statistical analyses.

Participants

A total of 342 students participated in this study; 4 of the students completed papers with content that could not be transcribed due to illegible handwriting. The final revision sample, therefore, consisted of 167 papers completed by Grade 5 students and 171 papers completed by Grade 8 students. Each student wrote two drafts, for a total of 676 writing samples. Although 40 fifth-grade students had IEPs, only those students whose primary disability was learning disabled or mildly mentally retarded and who had a writing goal on their IEPs were designated as special education for the purpose of this study (n = 24). The writing of other students with an IEP in speech, emotional disturbance, or hearing impaired was included with the larger sample (hereinafter labeled as the "general population"). The same was true for students in Grade 8; 46 students in the sample had IEPs, but only 29 met the criteria described previously. In an attempt to validate this grouping structure, we calculated effect size (ES) differences between the groups. A post hoc comparison of scale scores on the extended writing portion of the state test revealed smaller differences between the mean of students in general education and the mean of those students on IEPs without writing goals (ES = .07) and mean differences between this group and the group of students included in the special education sample (ES = .24). See Table 1 for student demographics.

State Assessment

Students completing the extended writing section of this state test were given 50 min during the first session to plan and write a first draft. During the second session, students were provided with an additional 50 min to revise, edit, and rewrite their compositions. Students were not given a choice of writing prompt, nor were they allowed to collaborate with peers during the revision session. The state department of education released the Grade 5 prompt; it read, "Imagine that you have become a hero for a day. Write a story about the day you became a hero. Remember to include characters, details, and a beginning, a middle, and an end." The prompt used at Grade 8 was not released;

thus it is not shared here. Scores on extended writing samples were reported as proficiency levels and scale scores. Scale scores were used in correlation analyses.

Revision Taxonomy

We used Faigley and Witte's (1981) revision taxonomy to guide our analysis of the quantity and quality of revisions made by fifth- and eighth-grade students on the state test. Although fairly complicated, a simple description of the revision taxonomy includes two primary categories: (a) *surface changes* and (b) *meaning changes*. Revisions classified as surface changes are further divided into two subcategories: *formal changes* (e.g., spelling, tense, and punctuation) and *meaning-preserving changes* (e.g., additions, substitutions, and deletions), none of which alter the message. Revisions classified as meaning changes are also divided into two subcategories: *microstructure changes* (e.g., additions, substitutions, and deletions resulting in minor changes to existing text) and *macrostructure changes* (e.g., additions, substitutions, and deletions constituting a major change to existing text).

Faigley and Witte's revision taxonomy, introduced in the early 1980s, highlighted the importance of categorizing revisions by their impact on the meaning of the writing, but, as described previously, it was a fairly complex model consisting of numerous gradations. Moreover, in the Faigley and Witte taxonomy, less emphasis was given to the specific unit of revision and more emphasis was placed on the depth (i.e., surface, meaning level) of the change (Faigley & Witte, 1981). In an attempt to collect additional information on revision types, Daiute's (1985) taxonomy included more emphasis on units of change (i.e., word, sentence, and intrasentence changes). MacArthur et al. (1991) added another distinction by rating each revision for quality, indicating whether or not specific revisions were better, worse, or neutral in comparison to the original text. In an attempt to create a comprehensive (but simple) taxonomy, we combined elements from all three of these taxonomies, while creating a total of only three exclusive and easily distinguishable categories: *unit, type*, and *quality*. See Figure 1 for a description of each level of the taxonomy. As the taxonomy makes clear, revisions are categorized first by *unit* (e.g., word, phrase, sentence), then by *type* (e.g., addition, substitution, spelling), and then by *quality* (i.e., whether the revision had an increased, decreased, or neutral effect on the text in comparison to the first draft.

Finally, a small distinction between our work and the work of other researchers is that we defined *unit* as a noun, whereas Graham (1997) labeled it a *syntactic unit*. In doing so, he included spelling under the unit heading, whereas we included spelling under *type*, which included any action students performed when making a revision (e.g., changing the spelling or substituting a phrase).

Interrater Agreement

We conducted three rounds of training to reach an acceptable level of agreement across the four raters. At the first training session the scoring rules were discussed and questions were clarified. Then, four raters (all professional educators) scored and discussed seven student compositions, one paper at a time. Minor inconsistencies surfaced across raters, and thus slight revisions were made to the scoring rule sheet.

Raters then read and coded 10 papers on their own, and initial reliability data were calculated using an agreements/agreements plus disagreements formula. Multiple raters and multiple papers were involved, and thus agreement on each paper was calculated in the following manner: Percentage of agreement between Raters 1 and 2, Raters 1 and 3, and Raters 1 and 4. Percentages for each paper were added and then divided by the number of pairs (three) for a total interrater agreement score. We applied this formula to each paper and then added percentages and divided by 10 for an overall average interrater agreement. Overall agreement was calculated at 81%. Data from this reliability analysis (and the qualitative information provided through raters' notes and questions) resulted in additions and clarifications to the set of scoring rules (see Figure 2).

A third and final round of training was held using the expanded set of scoring rules (see Figures 1 and 2). A batch of 47 student compositions was distributed to each reader, with 10 compositions duplicated across each batch. Agreement was quite strong across these 10 papers (95%), and thus it was determined that the changes made to the scoring rules and the last round of training resulted in strong enough agreement to score the final batch of papers without conducting another training session.

Analysis Procedures

Descriptive analyses were conducted to investigate the types of revisions students made and the differences in the types of revisions made across grade level and educational classification. We also investigated the relationship between the number of revisions made per 100 words and scores on the extended writing portion of the state test through a Pearson *r* correlation analysis. Finally, we analyzed qualitatively the types of revisions students made and explored patterns across students at different grade levels.

Figure 1
Revision Taxonomy

***Type* of Revision**	**Definition**
Deletion	Any unit may be deleted.
Substitution	Most often punctuation and words will be substituted.
Rearrangement	Phrases, sentences, and paragraphs will be rearranged. Changes usually improve style.
Spelling	Spelling changes occur at the word level.
Consolidation	Consolidations occur at the phrase, sentence, or paragraph level. A consolidation condenses multiple units into one cohesive unit.
Addition	Any unit may be added.
Expansion	Expansions occur at the phrase or sentence level. An expansion creates multiple units (e.g., sentences) where there once was one. Example: *I figured after walking so far the least he could do would be to provide a relaxing dinner since I was hungry.* *I figured the least he owed me was a good meal. All that walking made me hungry.*
***Unit* of Revision**	
Punctuation	Includes beginning, middle, and ending punctuation, and correct punctuation of abbreviations and titles.
Word	Includes proper nouns, abbreviations, etc.
Phrase	As few as two words should be scored as a "phrase."
Sentence	Must contain subject and predicate; correct punctuation unnecessary.
Multi-sentence	A group of sentences occurring within one paragraph.
Paragraph	A stand-alone paragraph.
Overall Judgment	
Surface	Basic editing or revisions of units that may or may not affect meaning within isolated occurrences.
Superficial	Unit changes occurring with enough breadth and/or depth to have some impact on meaning of passage (usually improving overall readability).
Substantive	"Macrostructural" revisions that function to change the meaning of the text (story) as a whole.

Source: Adapted from "Analyzing Revision," by L. Faigley and S. Witte, 1981, *College Composition and Communication*, *32*, p. 400–414.

Results

Unit, Type, and Quality Analyses

In terms of unit changes, word revisions were the most prevalent (approximately 40%) across all students in Grades 5 and 8. Approximately 25% of revisions were phrase changes, defined as two or more words strung together, with students in Grade 8 making slightly more phrase changes than students in Grade 5 (see Figure 3). The unit categories with the fewest number of revisions included larger sections of text (paragraph, multisentence, and multiparagraph revisions), across both grades.

When coding for type of revision, substitutions accounted for approximately 45% of all revisions, with no pattern across the different subgroups. Additions represented approximately 20% of all type revisions, with students in the general population making slightly more additions than students in special education across both grade levels. The reverse pattern was true for spelling revisions, with students in special education making more spelling revisions than their counterparts in the general population (16% vs. 10%; see Figure 4).

Little variability across the four groups of students was apparent in terms of the quality of individual revisions. Approximately 60% of all revisions were coded as increasing quality, and approximately 20% were coded as decreasing quality; the remaining revisions were scored as "neutral" (see Figure 5).

Figure 2
Additional Scoring Rules

Two or more paragraphs are created when there once was one:	*Rearrangement*
One paragraph combined with another paragraph:	*Consolidation*
In either of the above situations, continue to code revisions within the paragraph.	

Phrase to word	*Word substitution*
Word to phrase	*Phrase substitution*
Both of the above rules apply except when the revision in question is a contraction.	
In either direction, changes involving contractions:	*Word substitution*
Similarly, when a change of tense occurs (e.g., "be" to "has been"):	*Word substitution*

More than one sentence rewritten as one sentence:	*Multi-sentence consolidation*

Changes that appear to be spelling errors (e.g., "I **am** going" to "I **a** going"):	*Word substitution*
Singular to plural or plural to singular (e.g., "men" to "man," or "I" to "we"):	*Word substitution*
Added or deleted suffixes or prefixes:	*Word spelling*

When word is deleted from rough draft: "**There i**n front," to "**I**n front," making it necessary to capitalize **I**:	*Word deletion - not punctuation addition*

Two or more phrase changes (e.g., substitution, redistribution, etc.):	*Sentence level*

Type by Quality

These findings prompted the following question: What was the relationship between type of revision and quality of revision? Tables 2 and 3 provide data related to the type of revisions most apt to be judged as an improvement in quality over the first draft. The tables highlight the positive impact of additions to students' final drafts. For students without IEP goals in writing, additions were rated as an improvement 76% of the time at Grade 5 and 70% of the time at Grade 8. For students with IEP goals in writing, additions were coded as an improvement 66% and 67% of the time in Grades 5 and 8, respectively. Deletions had the least positive impact across categories, grades, and educational classifications, resulting in an increased quality score only 42% of the time, with students in fifth grade gaining the least from deletions (i.e., resulting in an increased quality score only 3.5 times per every 10 deletions). The relationships between other types of revisions and their impact on quality also are presented in Tables 2 and 3.

Quantity and Overall Impact of Revisions

Before calculating total number of revisions made, we calculated total number of words written on first and final drafts and reported these numbers across grade level and educational classification. We then calculated number of revisions made and finally, number of revisions per 100 words (see Table 4). The quantity of revisions revealed some patterns, with students in the general population making almost twice as many revisions as their counterparts in special education. It is important to note, however, that students in the general population also wrote almost 100 words more than students in special education at Grade 5. At Grade 8, students in the general population wrote over 150 more words than their counterparts in special education. Quantity of revisions was thus confounded by total number of words written, as was made obvious in the calculation of revisions per 100 words. Once the effect of rate of writing was neutralized, minimal differences existed in the number of revisions made across students in the general population and

Figure 3
Percentage of Revisions by Unit

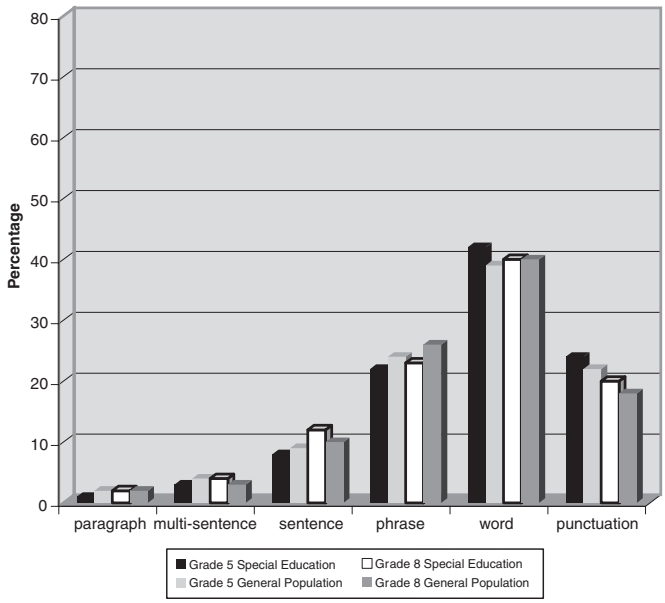


Figure 5
Percentage of Revisions by Quality

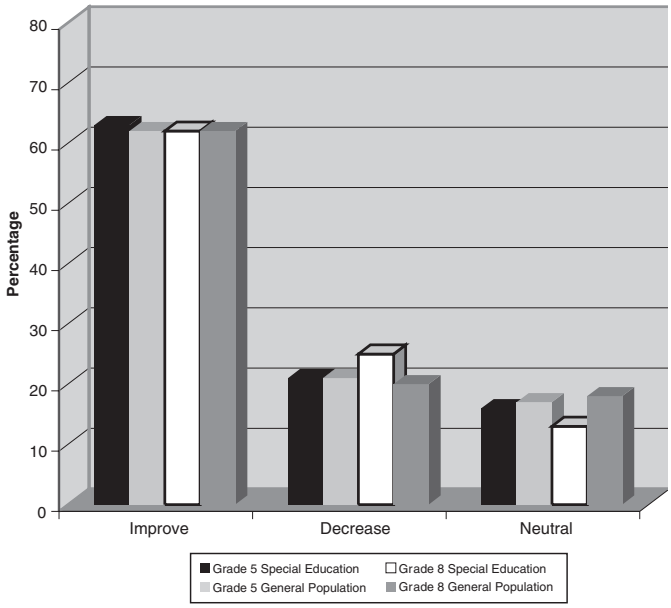
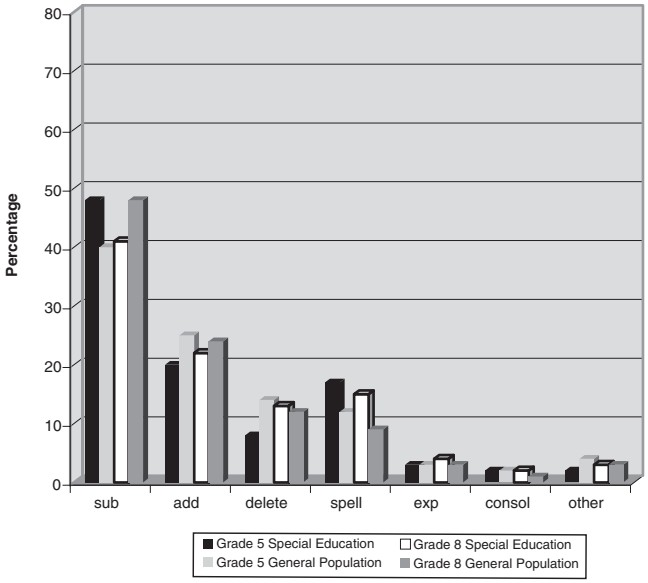


Figure 4
Percentage of Revisions by Type



students in special education who had IEP goals in writing (see Table 4).

As a final series of analyses, overall impact of revisions was explored through the relationship between the number of revisions made per 100 words and students' earned scale score for the extended writing sample on the state test. At Grade 5, no relationship existed between these two variables, $r(166) = -.07$, while a moderate correlation was found between total words written and score on the extended writing sample, $r(166) = .54$. Similar results were found for Grade 8; the correlation between revisions per 100 words and state test score on the extended writing sample was weak and negative, $r(151) = -.21$, while the correlation for total words written and extended sample was much stronger, $r(151) = .68$. Even though correlations were negative for revisions per 100 words and scale score on the extended writing sample (i.e., the more revisions made, the lower the score), the weak relationship between the two variables negates the impact of this finding. Owing to the small number of participants in special education, we did not disaggregate the data used for correlation analyses, but instead used aggregate data sets, including all students, at Grades 5 and 8. Finally, the state test data received at Grade 8 was missing scores for 19 students.

Qualitative Observations

As the four raters submitted their scored samples, they reported that for the most part students' final drafts read qualitatively better than their first drafts. Raters reported that even when students made revisions that decreased the quality of individual units, they usually made enough positive revisions to improve overall readability of the composition. Qualitative findings supported quantitative

Table 2
Relationship Between Type and Quality of Revisions: Grade 5

	Addition, *n* (%)	Deletion, *n* (%)	Substitution, *n* (%)	Spelling, *n* (%)
Increase				
General population	874 (76)	209 (33)	1,129 (61)	344 (62)
Special education	59 (67)	14 (38)	139 (67)	41 (54)
Neutral				
General population	157 (13)	152 (24)	376 (20)	40 (7)
Special education	10 (11)	10 (27)	38 (18)	7 (9)
Decrease				
General population	132 (11)	267 (43)	336 (18)	170 (31)
Special education	19 (22)	13 (35)	32 (15)	28 (37)
Total				
General population	1,163 (100)	628 (100)	1,841 (99)	554 (100)
Special education	88 (100)	37 (100)	209 (100)	76 (100)

Note: Individual cells represent the total number of revisions made by fifth graders and the percentage of those types of revisions that increased, decreased, or had a neutral effect on quality of text.

Table 3
Relationship Between Type and Quality of Revisions: Grade 8

	Addition, *n* (%)	Deletion, *n* (%)	Substitution, *n* (%)	Spelling, *n* (%)
Increase				
General population	698 (70)	273 (53)	1,190 (61)	234 (61)
Special education	106 (66)	43 (45)	185 (61)	67 (61)
Neutral				
General population	164 (16)	120 (23)	381 (19)	21 (6)
Special education	19 (12)	21 (22)	53 (17)	7 (6)
Decrease				
General population	135 (14)	127 (24)	395 (20)	126 (33)
Special education	36 (22)	32 (33)	65 (21)	35 (32)
Total				
General population	997 (100)	520 (100)	1,966 (100)	381 (100)
Special education	161 (100)	96 (100)	303 (99)	109 (99)

Note: Individual cells represent the total number of revisions made by eighth graders and the percentage of those types of revisions that increased, decreased, or had a neutral effect on quality of text.

Table 4
Average Total Words Written and Revisions per 100 Words: Grades 5 and 8

	First Draft Total Words	Final Draft Total Words	Average Number of Total Revisions	Revisions per 100 Words	*SD*
Grade 5					
General population	271	281	32	11	6.8
Special education	184	201	18	10	7.2
Grade 8					
General population	289	305	32	11	5.0
Special education	138	142	17	12	8.7

findings to a large extent. For example, approximately one third of all spelling substitutions were scored negatively because students changed the spelling of a word in the final draft but still spelled it incorrectly. Also, additions almost always improved the piece, and finally, some students made revisions perceived as "safe" by the scoring team. We share examples of these three patterns in the paragraphs that follow.

Students made numerous spelling revisions in their final drafts, but almost 40% of the time, these changes were rated as "decrease" or "neutral." For example, one student changed the correct spelling of *robbery* in the first draft to the incorrect spelling of *robbary* in the final draft; another student substituted *sommone* in the final draft for *somone* in the first draft, and a different student substituted *dinoser* for *dinosor*. The number of these types of revisions makes it hard to share the range of examples, but these examples are typical of those made by students. It also was not unusual to read stories in which students substituted words in an attempt to use correct spelling; yet, in many cases, the new words were also spelled incorrectly. These changes were coded as "word substitutions," but raters perceived them as being driven by a concern for spelling of the initial word. For example, substitutions such as "stabed me" for "knived me," or in a different type of error, the spelling of the possessive "thier" to "there" in the final draft.

On the other hand, raters often shared that additions seemed to improve the stories, and in fact, additions were rated as improvements more than 65% of the time across both grade and educational classifications. For example, in a first draft, a student wrote, "Outside was a mid evil town!" and in the second draft the student revised the sentence to read, "I slowly opened the door and saw a mid evil town full of people!" Oftentimes, additions came at the beginning or end of a story, providing the reader with a better introductory set or conclusion. Raters commented, though, that sometimes these additions felt staged, as if students knew they would need to show revisions in their final drafts so they purposefully did not develop these sections in their first drafts; for example, one student concluded his story with, "I thought sometimes a fire can save your life other times it is a huge yellow, orrnge and red deth trap." In his final draft he concluded with, "I thought sometimes a fire can save your life other times it's a huge yellow, orenge, and red deth trap. . . . I was reworeded with a meadel and a cash reward. That ment my mom only had to work one job, my sister could stay in colage and I was never called a kid agne!" We could not, however, validate the observation that students intentionally underdeveloped introductions and conclusions in their first drafts.

Finally, at times it appeared that students substituted words or phrases that had less voice than in their first drafts or made substitutions in an attempt to improve their spelling. In these cases, substitutions did not always improve the piece. For example, in his first draft, one student wrote, "[The man] was sentenced to 10 years in prison for murder, auto theft, burgaleries, bank robbing attempts, and snipering!" His seemingly safe revision in the final draft read, "[The man] was sentenced to twenty years in prison for many serious crimes." The revised sentence was spelled correctly but was void of the voice present in the first sentence.

Discussion

The impetus for this study was the limited amount of research into the revision behaviors of students on state tests. A great deal of research has been conducted in controlled or classroom settings, but it is not known if findings from these studies generalize to state assessment contexts. We did, in fact, uncover similarities between revision behaviors of students in assessment situations and previous intervention studies, but not all of the results reported have been previously documented in the literature.

Our revision analyses were structured around three primary research questions. First, we explored the kind of revisions that students made on state assessments. Findings mirror those from previous research in that students made an overwhelming number of word and phrase revisions and far fewer sentence and paragraph revisions. We also found that students substituted and added text more often than they deleted text and that additions were generally positive, while deletions were more often negative. Across all students, revisions were rated as improvements at least 65% of the time.

The second research question related to possible differences in revision behaviors across students in the Grades 5 and 8 and students with and without disabilities. Somewhat unexpectedly, we found few differences across students at different grade levels and across students with and without IEPs in writing in terms of the type, unit, and quality of revisions; yet, researchers have reported differences in the quality of revisions between young and more experienced writers (Butterfield, Hacker, & Plumb, 1994) as well as the quantity of revisions between younger and older writers (see review by Fitzgerald, 1987). Furthermore, students in special education have been found to struggle with executive control (Graham, 1997; Scardamalia & Bereiter, 1983) and self-regulatory strategies (Graham & Harris, 1994) that would enable them to

revise effectively. We found, however, that once rate of writing was eliminated from the analysis, students in special education made a similar number of overall revisions as students in the general population.

Our third and final question explored the relationship between the number of revisions students made and their scores on the extended writing portion of the state test. Interestingly, no correlation between these two variables existed, but a moderate correlation did exist between total words written and state test scores across both grade levels. In no way does this finding imply causation, but it does demonstrate a relationship between rate of writing and score received on a state writing exam.

Instructional and Assessment Implications

The descriptive data reported in this study highlighted students' tendencies to make surface-level revisions consisting primarily of word, phrase, and punctuation changes. Across all students, our results parallel those reported by other researchers in that "form received more emphasis than substance" (Graham, 1997, p. 232). Descriptive data support this finding, as do the qualitative observations of students' choice to make "safe" revisions that improved form but did not necessarily improve content, and many times decreased the level of voice in their writing.

Student focus on form over content is disconcerting to many professionals in the field and may become an increasing concern as we continue to align our classroom instruction too closely with the limited demands of our large-scale assessments of writing, which cannot focus too heavily on substance due to time constraints and the need to achieve interrater reliability during scoring sessions. As George Hillocks (2002) expressed,

> States are investing millions of dollars, thousands of teacher hours, and hundreds of thousands of student classroom hours in mandatory writing assessments. If the investment of all this money and time is having a positive effect on student achievement, it may be worthwhile. If assessments limit the kinds of writing taught or the ways they are taught, or the thinking that good writing requires, then the assessments may be of questionable value. (p. 19)

Concern about the influence of state tests on classroom instruction may be valid. For example, Clare, Valdes, and Patthey-Chavez (2000) found that the feedback students received during classroom writing assignments tended to focus more on the mechanical aspects of writing than on content, organization, and more substantive features of composition. In a follow-up study, these findings generalized across teachers working in elementary schools with low-, middle-, and high-achieving writers. Teachers' feedback to students was described primarily as helping students "standardize" or edit the writing produced in their rough drafts, as opposed to helping students elaborate on or improve their overall compositions (Matsumura et al., 2002). An alternative explanation for students' reliance on low-level revisions may be "an unwilling[ness] to take risks in their writing for fear of failure" (Troia & Maddox, 2004, p. 32). Fear of failure is a valid concern for all students, but perhaps even more so in the context of high-stakes state assessments.

A more straightforward instructional implication from this study is that teachers may want to instruct students to add or substitute information in their final drafts but delete text only when absolutely necessary. Also, it may be wise to teach students to change the spelling of a word in their final drafts only if they are sure of its correct spelling. If not, the data reported here suggest that students should maintain the original spelling of a word.

Perhaps the most important finding is the relationship between rate of writing and score on the extended writing test. Moderate correlations have been reported between total words written and scores on more global measures of writing (Moss, Cole, & Khampalikit, 1982; Tindal & Parker, 1989), but these analyses were not conducted within the context of state assessments. One possible instructional implication is that students in special education must learn to write more fluently. Among other things, scoring rubrics evaluate content and organization. Without a reasonable number of words, the content of a story is limited, and organizational skills cannot be easily demonstrated. It also seems reasonable to assume that evaluators will score negatively other constructs, such as style, fluency, and word choice, when presented with a paucity of text.

Finally, at the risk of overstating these findings, we found ourselves asking some questions about the relationship between teaching and testing. For example, Are teachers becoming too formulaic in teaching the writing process? How has the structure of state tests impacted instruction, both positively and negatively? What are we teaching students about how and when to revise? A follow-up question has to do with the emphasis on revision for the sake of revision. As Wolf (1993) stated, "Failing to come up with radical and plentiful revisions to a first draft is to fall short of contemporary writing process expectations that everyone is an author and that authorship entails revision" (p. 214); yet, the majority of revisions students made on this state test were of limited quality, and the quantity of revisions they made had no relationship to the score they earned.

References

Beal, C. R. (1987). Repairing the message: Children's monitoring and revision skills. *Child Development*, *58*, 401–408.

Beal, C. R. (1990). The development of text evaluation and revision skills. *Child Development*, *61*, 247–258.

Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Topics in Language Disorders*, *20*(4), 65–84.

Bridwell, L. S. (1980). Revising strategies in twelfth-grade students' transactional writing. *Research in the Teaching of English*, *14*, 197–222.

Butterfield, E. C., Hacker, D. J., & Plumb, C. (1994). Topic knowledge, linguistic knowledge, and revision processes as determinants of text revision. In J. S. Carlson (Series Ed.) & E. C. Butterfield (Vol. Ed.), *Advances in cognition and educational practice: Vol. 2. Children's writing: Toward a process theory of the development of skilled writing* (pp. 83–143). Greenwich, CT: JAI Press.

Clare, L., Valdes, R., & Patthey-Chavez, G. G. (2000). *Learning to write in urban elementary and middle schools: An investigation of teachers' written feedback on student compositions*. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Colorado Department of Education. (2005). *Fact sheet for writing: Unit for student assessment*. Retrieved October 1, 2006, from http://www.cde.state.co.us/cdeassess/documents/csap/frameworks/CSAP_factsheet_06_wrtg.pdf

Daiute, C. (1985). Do writers talk to themselves? In S. W. Freedman (Ed.), *The acquisition of written language: Response and revision* (pp. 133–159). Norwood, NJ: Ablex.

Daiute, C., & Kruidenier, J. (1985). Special issue: The psycholinguistics of writing. *Applied Psycholinguistics*, *6*, 307–318.

De La Paz, S., & Graham, S. (1997). Strategy instruction in planning: Effects on the writing performance and behavior of students with learning difficulties. *Exceptional Children*, *63*, 167–181.

Delaware Department of Education. (2006). *Delaware student testing program: Frequently asked questions*. Retrieved October 1, 2006, from http://www.doe.k12.de.us/aab/FAQ_2006.pdf

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, *32*, 400–414.

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research*, *57*, 481–506.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365–387.

Flower, L., Hayes, J. R., Carey, L., Shriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies of revision. *College Composition and Communication*, *37*, 16–55.

Gersten, R., & Baker, S. (2001). Teaching expressive writing to students with learning disabilities: A meta-analysis. *Elementary School Journal*, *101*(3), 251–272.

Goldberg, G. L., Roswell, B. S., & Michaels, H. (1995–1996). Can assessment mirror instruction? A look at peer response and revision in a large-scale writing test. *Educational Assessment*, *3*, 287–314.

Graham, S. (1997). Executive control in the revising of students with learning and writing difficulties. *Journal of Educational Psychology*, *89*, 223–234.

Graham, S., & Harris, K. R. (1994). The role of self-regulation in the writing process. In D. Schunk & B. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 203–228). Hillsdale, NJ: Lawrence Erlbaum.

Graham, S., MacArthur, C., & Schwartz, S. (1995). Effects of goal setting and procedural facilitation on the revising behavior and writing performance of students with writing and learning problems. *Journal of Educational Psychology*, *87*, 230–240.

Hillocks, G., Jr. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.

MacArthur, C., & Graham, S. (1987). LD students composing under three methods of text production: Handwriting, word processing, and dictation. *Journal of Special Education*, *21*, 22–42.

MacArthur, C. A., Graham, S., & Schwartz, S. (1991). Knowledge of revision and revising behavior among students with learning disabilities. *Learning Disability Quarterly*, *14*, 61–73.

Matsumura, L. C., Patthey-Chavez, G. G., Valdés, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *Elementary School Journal*, *103*, 1–25.

Moss, P. A., Cole, N. S., & Khampalikit, C. (1982). A comparison of procedures to assess written language skills at grades 4, 7, and 10. *Journal of Educational Measurement*, *19*, 37–47.

National Assessment of Educational Progress. (1986). *The writing report card: Writing achievement in American schools*. Princeton, NJ: Author.

Perez, S. A. (2000). Revising during writing in a second grade classroom. *Educational Research Quarterly*, *25*(1), 27–32.

Rohman, D. G., & Wlecke, A. O. (1964). *Pre-writing: The construction and application of models for concept formation in writing*. East Lansing: Michigan State University.

Scardamalia, M., & Bereiter, C. (1983). The development of evaluative, diagnostic and remedial capabilities in children's composing. In M. Marlew (Ed.), *Psychology of written language: A developmental and educational perspective* (pp. 67–96). New York: John Wiley.

Stoddard, B., & MacArthur, C. A. (1993). A peer editor strategy: Guiding learning-disabled students in response and revision. *Research in the Teaching of English*, *27*, 76–103.

Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *Journal of Special Education*, *23*, 169–183.

Troia, G. A., & Maddox, M. E. (2004). Writing instruction in middle schools: Special and general education teachers share their views and voice their concerns. *Exceptionality*, *12*(1), 19–37.

Whitaker, D., Berninger, V., Johnston, J., & Swanson, L. (1994). Intraindividual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences*, *6*, 107–130.

Wolf, D. P. (1993). Assessment as an episode of learning. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 213–240). Hillsdale, NJ: Lawrence Erlbaum.

Wong, B. Y. L., Butler, D. L., Ficzere, S. A., & Kuperis, S. (1996). Teaching low achievers and students with learning disabilities to plan, write, and revise opinion essays. *Journal of Learning Disabilities, 29,* 197–212.

Lindy Crawford, PhD, is an assistant professor in and chair of the department of Special Education at the University of Colorado at Colorado Springs. Dr. Crawford's current research

interests include analyzing student performance on large-scale writing assessments, the meaningful inclusion of students with disabilities in statewide test programs, and effective instructional strategies for English language learners.

Susan Lloyd, PhD, is a reading specialist in Fairfax County, Virginia. Dr. Lloyd's current research interests include implementation of guided reading instruction, writing conferences with students, and schoolwide change based on instructional walk-through data.

Kelly Knoth recently graduated from the University of Colorado at Colorado Springs with an MA in curriculum and instruction with an emphasis in reading.