

Effects of NWP Teaching Strategies on Elementary Students' Writing

Sherry Seale Swain, Richard L. Graves, and David T. Morse

National Writing Project

July 2007

Table of Contents

Executive Summary	4
Introduction	5
The Professional Development Program	5
Program Background and Description.....	6
Intended Influence on Student Outcomes	7
Foci of Workshops and Classroom Demonstrations.....	7
Factors Affecting Program Design	8
Research Questions	8
Questions Related to Teacher Practices	8
Questions Related to Student Outcomes.....	9
Questions Related to Correlations Between Measures	9
Research Design and Methods	9
Overall Research Design.....	9
Sample.....	9
Characteristics of Program and Comparison Schools.....	9
Teacher Populations.....	11
Data Collection	11
Variables	11
Data Sources	11
Collection of Teacher and Student Data	13
Data Analysis	14
Framework for Evaluating and Scoring Student Writing	14
Rubric.....	14
Scoring	15
Syntactic and Prominent-Feature Analysis.....	15
Selection of the Research Team.....	15
Prominent-Feature Analysis.....	16
Developing the Prominent-Feature Score.....	18
Syntactic Measures Analysis	18
Measures of Syntactic Maturity	18
Analysis of Teachers' Degree of Implementation	19
Results and Findings	19
The Effect of Professional Development on Teacher Practices.....	19
Use of Strategies	20
Sum of Strategies Score	22
Degree of Implementation	22
Student Outcomes	23
Student Writing Performance, Holistic and Analytic	23
Summary of One-Between, One-Within ANOVA Analyses	25
Student Writing Performance, Syntactic Measures, and Prominent Features	26
Frequency of T-units, Clauses, and Prominent Features	26
Mean T-Unit Length Comparisons	29

Correlation of Prominent-Feature Scores and Length Indicators	30
Pre and Post Frequencies of Prominent Features.....	31
Correlation of Prominent-Feature Scores with NWP Analytic, Holistic Scores	33
Teacher Practices and Student Outcomes	35
Conclusions and Discussion	36
Effect on Teaching Practices	37
Effect on Student Outcomes	37
Analytic Writing Scores and Student Outcomes	37
Syntactic Measures and Student Outcomes	38
Prominent Features and Student Outcomes	39
Implications for Professional Development.....	39
Bibliography/References	40
Appendix A. Interview Guide	43
Appendix B: Prompts	45
Appendix C: Definition of Strategies	46
Appendix D: Continuum of Implementation	47

Executive Summary

This study of elementary students' writing builds upon a 2005 study of seventh grade writing in which papers returned from a state assessment were analyzed to determine what, if any, prominent or syntactic features in the papers correlated to the holistic state scores. Results of that study indicated strong correlations between certain writing features and higher scores on the holistic state test and prompted the present study.

The present study employed a quasi-experimental design, assessing the writing of program and comparison groups of upper elementary students using pre and post writing assessments. Two similar schools participated in the study; only the program school was involved in writing project professional development. The sample included 26 teachers and 435 students in the program school, and 16 teachers and 217 students in the comparison school.

Research questions focused on the differences in growth between pre and post assessments in the writing of third, fourth, and fifth grade students, some of whom were taught by teachers involved in writing project professional development and some by teachers who were not involved. Student papers were holistically and analytically scored at a national scoring conference sponsored by the National Writing Project. Teaching practices of faculties in both schools were analyzed to determine which of the strategies advocated in the professional development were ultimately implemented in classrooms and for the individual degree of implementation for each teacher. Additionally, the study examined prominent and syntactic feature scores on pre and post assessments, and the correlations among the feature scores and the NWP writing assessment scores. In order to better understand the specific features that appear in higher-scoring papers, this study also examined correlations of the individual prominent features found in student writing to the scores derived from the NWP Analytic Writing Continuum.

Student growth in writing between pre and post assessments was significantly greater for the program students. Even though comparison students scored higher than program students on the pre assessment, program students significantly outperformed comparison students on the post assessment—the holistic score as well as each of the analytic scores for content, structure, stance, sentence fluency, diction, and conventions. Findings indicate that the teaching practices of teachers participating in the writing project professional development were more process-based and student-centered than those of teachers in the comparison group; teachers with the highest implementation of the NWP teaching strategies fostered the greatest growth in student writing. Findings further show statistically significant correlations between a number of the prominent features and the analytic writing traits. All in all, this study has much to say about the teaching of writing—what to teach and how to teach it—and about professional development for teachers of writing.

Introduction

The Mississippi Writing/Thinking Institute (MWTI), comprising the seven National Writing Project (NWP) sites in Mississippi, is housed in the College of Education at Mississippi State University. Through its seven university-based NWP sites, MWTI aims to improve the teaching of writing throughout the state. MWTI and its local sites provide intensive, multiyear support for schools and districts; this support includes school-based professional development workshops, classroom demonstrations, study groups, and coaching and mentoring for individual teachers.

In an earlier study of student writing, MWTI researchers analyzed state-scored seventh grade papers and identified a number of syntactic and prominent rhetorical features that correlated highly to better scores on a holistic state assessment (Swain, Graves, and Morse 2006). In designing the present study, researchers wondered whether similar correlations might result from a prominent-feature analysis of papers scored using the NWP Analytic Writing Continuum assessment system, a system designed to measure growth in student writing at various grade levels from sites across the nation. If so, MWTI researchers predicted that a prominent-feature analysis of student writing could positively influence subsequent professional development offered to teachers. Therefore the professional development for the present study utilized a prominent-feature analysis of a pre assessment of students' writing to influence content of professional development offered in the program school.

The present study compares teacher practices and student outcomes for students in grades 3, 4, and 5 in two elementary schools, one of which participated in MWTI professional development during the 2005–2006 school year. The study focuses on differences between and within pre and post writing assessments on papers from the program and comparison schools, as measured by the NWP Analytic Writing Continuum and holistic rubric. It also focuses on differences between and within the pre and post writing assessments on syntactic and prominent-feature measures of all papers from the program school and a random sample of papers from the comparison school. In addition, teacher practices are analyzed and compared with student outcomes.

The Professional Development Program

Program Context

The study focuses on two elementary schools, each serving third-, fourth-, and fifth-graders. Both schools have achieved high ratings on the state accreditation scale; both serve populations of primarily Caucasian students, with small populations of African American students, and even smaller populations of newly arrived Hispanic immigrants. The program school is located in a rural area near Mobile, Alabama, with 53% of its 593 students receiving free or reduced-price lunches; the comparison school is in a suburban area near Memphis, Tennessee, with 59% of its 348 students receiving free or reduced-price lunches. The program school, located some 20 miles inland from the Gulf Coast, was greatly affected by Hurricane Katrina at the beginning of the study, with children's and teachers' families and lives interrupted in varying degrees throughout the year.

Program Background and Description

In the summer of 2005, the principal of the program school approached MWTI staff to request an initial year of professional development in writing instruction for teachers in grades 3, 4, and 5 in her upper elementary school, a school that had already achieved the highest accreditation rating awarded by the state. Explaining that the accreditation was based only on reading and mathematics, the principal noted that writing was a weak area and that the teachers had never had professional development focused on writing. MWTI worked with the principal to design a professional development program that would introduce teachers to writing-process teaching strategies, and would fit within the scheduling constraints of the school day.

The following two charts show the major activities of the professional development at the program school, both on and off site.

MAJOR ON-SITE ACTIVITIES AND FEATURES

On-Site Structures	Description	Number of Sessions	Number of Hours
Half-day workshops	Interactive sessions in which participants personally experience effective practices tied to learning theories and then reflect on how those practices can be implemented in their own classrooms.	7	21
Preparation for classroom demonstrations (1 hour each)	Presession meetings in which teachers are prepared to observe and participate in classroom demonstrations, including the rationale for the lesson to be presented and the rationale for the processes used in the presentation.	7	7
Classroom demonstration days	Three to four classroom demonstrations are presented per day. Small groups of teachers participate in each demonstration, writing with the students, assisting small groups of students with assignments, reflecting afterward on their insights related to the lesson.	12 days (3–4 demonstrations per day) 40 individual demonstrations	
Classroom demonstrations per teacher (1 hour each)	Each classroom hosts a classroom demonstration at least one time so that all students have the opportunity to be “taught” for one class by an MWTI staff developer and a small group of teachers from other classrooms. Each individual teacher attends 6 such sessions.	6	6
Total Program Time per Teacher			34 hours

MAJOR OFF-SITE ACTIVITIES AND FEATURES

Off-Site Structures	Description
Lesson plans	MWTI staff developers provide lesson plans so that teachers can replicate and adapt the classroom demonstration lessons. Lesson plans include suggestions for the use of large-group, small-group, and individual instruction; guides for moving through the process focus of the lesson; and a rationale explaining how each part of the process fits into the learning schema. Teachers are encouraged to adapt the lessons for their particular students.
Ongoing multiple drafts	MWTI staff developers ask that teachers work on a single piece with each student, taking that piece through multiple drafts rather than assigning new writing prompts daily.
Model responses to student writing	MWTI staff developers work with the principal to establish dates by which students should complete new drafts of their pieces. These pieces are then shipped to the MWTI office where staff developers write a response to each student, providing a model for teachers—a scaffold for teachers to use as they learn how to respond to student writing in ways that lead to work writing improvement.

Intended Influence on Student Outcomes

In early November, research team members read the pre assessments and debriefed their impressions of the prominent features in the writing with MWTI staff developers (the statistical analysis of the prominent features appearing in the pre writing assessments was not formally completed until after the post writing assessment had occurred). These observations influenced the professional development program intended to positively impact students' writing.

Foci of Workshops and Classroom Demonstrations

In addition to presenting writing process approaches appropriate for initial stages of staff development, MWTI staff developers included topics suggested by the research team—suggestions gleaned from informal impressions of prominent features on the pre assessment. The resulting program, then, was not totally based on a formal analysis, but was influenced by the research team's initial observations of the prominent features exhibited in the pre writing.

Specific topics for workshops and classroom demonstrations included

- Introduction to writing process
- Writing process
- Student choice
- Vivid verbs

Show writing
Idea generation
Full description
Attention to detail
Analysis of first-draft writing
Organization
Conclusions
Dialogue
Cumulative sentences
Adverbial leads
Multiple revision strategies
Writing on demand
Teacher/student conferences and appropriate questioning strategies
Student/teacher reflection on progress.

Factors Affecting Program Design

The design of the program—with the goal of influencing student writing—was based on three factors. First, the intensity of contact with the teachers (34 hours) was likely to create change (Swain, Graves, and Morse 2005a). Second, the content and format of the program was focused on measurable features of student writing. Previous studies have indicated that content focused on related areas such as writing across the curriculum do not influence, directly or indirectly, writing achievement (Swain, Graves, and Morse 2005b). Finally, the research team’s suggestions for program foci meant that staff developers did not rely on a decontextualized list of desired writing skills and abilities, but rather designed a program focused on observed student need.

Research Questions

Research questions focus on three areas: teacher practices; student outcomes; and correlations of syntactic measures and prominent features of student writing, with student outcomes. Questions related to *teacher practices* focus on the extent to which teachers in the program school adopted teaching practices from the staff development, which of those strategies appeared to be most and least readily adopted by teachers, and whether or not there were differences between practices of program and comparison teachers. Questions related to *student outcomes* focus on whether or not there are differences in growth of program and comparison groups on the pre and post writing assessments. Questions related to *syntactic measures and prominent features* of pre and post writing assessments for both groups focus on correlations with the assessment data.

Questions Related to Teacher Practices

1. To what extent, if at all, do teachers in their first year of MWTI professional development use the recommended strategies in their classrooms?
2. What differences, if any, exist in the teaching strategies of program versus comparison teachers?

Questions Related to Student Outcomes

3. What differences, if any, exist in scores of program versus comparison students on pre and post writing assessments from a national scoring conference using the NWP Analytic Writing Continuum and holistic rubric?
4. What syntactic features are present in the pre and post writing assessments of upper elementary students?
5. What prominent features are present in the pre and post writing assessments of upper elementary students?

Questions Related to Correlations Between Measures

6. What correlations, if any, exist between measures of syntactic maturity and score points on the NWP holistic rubric?
7. What correlations, if any, exist among the prominent features and the scores derived from the NWP Analytic Writing Continuum and holistic rubric?
8. What relationships, if any, exist among teacher practices and student outcomes?

Research Design and Methods

Overall Research Design

This study employed a pre and post quasi-experimental design. Because the study focused on a contracted program with a whole school, randomization was not possible. Similarities in demographic factors such as school accreditation level, type of community, percentage of free and reduced-price lunches, ethnic diversity, and per-pupil expenditures were used to identify a closely matched comparison school within the state. Data collected from both schools included teacher interviews and pre and post samples of student writing.

Sample

The program school was self-selected through a contract for professional development with MWTI and through agreement to participate in the study. The comparison school was selected based on its similar demographics and school accreditation level. The principal at the comparison school enthusiastically agreed to participate in the study in exchange for reports from the national scoring conference. At the conclusion of the study, both schools received the assessment information, and both principals stated their intent to use the information to improve the teaching of writing in their schools.

Characteristics of Program and Comparison Schools

Specific characteristics of the students in both schools are shown in table 1.

Table 1. Student Demographics, Program and Comparison Schools, 2005–2006

Criteria	Program School	Comparison School
Served by NWP Staff Development	Yes	No
Percentage of Students on Free/Reduced-price Lunch	53	59
Racial Diversity (percentage) (African American / Caucasian/Hispanic)	8/92/0	26/67/6
Third Grade Student Population	161	117
Fourth Grade Student Population	188	113
Fifth Grade Student Population	244	118
Expenditures per Pupil (2001–2002)	\$5,311	\$4,584

Source: Mississippi Public Schools Fall Enrollment, 2001–2005, Statistics Division, Office of Management Information Systems, Mississippi State Department of Education; Mississippi State Department of Education Website: www.mde.k12.ms.us/

Finally, data on language arts and reading scores confirmed the similarity of program and comparison schools, with percentages of program and comparison students scoring “basic and above” falling within a range of 9%. For each group of students, researchers examined language arts and reading scores on the spring 2005 Mississippi Curriculum Test (MCT), scores from the end of the year preceding this study. These data are presented in table 2.

Table 2. Percentages of Students, Grades 2, 3, 4, Scoring Basic and Above in Language Arts and Reading, Mississippi Curriculum Test (MCT)

Baseline Year	Program School Spring 2005*			Comparison School Spring 2005*		
	Grade 2	Grade 3	Grade 4	Grade 2	Grade 3	Grade 4
MCT Basic and Above in Language Arts	98.4	99.4	97.3	95.8	90.6	98.1
MCT Basic and Above in Reading	98.4	98.8	97.3	96.7	96.3	100
Writing Assessment Scores of 2 or Above**	-	-	94	-	-	96

Source: Mississippi State Department of Education Statistics Division, Office of Management Information Systems, Mississippi State Department of Education; Website: www.mde.k12.ms.us/

Note: *Grade 2 scores from 2005 are intended to represent Grade 3 students in the study, and so on.

**Mississippi Writing Assessment scores are measured by a 4-point holistic rubric.

Teacher Populations

Of the 26 program school teachers, 11 were interviewed for this research. Of the 16 comparison school teachers, 8 were interviewed. As shown in table 3, most comparison teachers had earned advanced degrees, while two program teachers had earned national board certification. No teachers were in their first year of teaching, and most teachers in both schools had six or more years of experience. During the year of the study (2005–2006), all program teachers were participating in their first year of MWTI professional development.

Table 3. Percentage of Interviewed Teachers' Educational Level and Level of Experience for Program and Comparison Schools

	Certification Level			Years of Experience				
	B.S.	M.Ed.	National Board Certification	1	2–5	6–10	11–15	16+
Program School <i>N</i> = 11	73%	27%	18%	0%	9%	36%	27%	27%
Comparison Schools <i>N</i> = 8	25%	75%	0%	0%	13%	50%	0%	37%

Source: Teacher interviews.

Data Collection

Variables

Independent variables related to measures of teacher practices included 1) participation or nonparticipation in MWTI professional development, and 2) levels of education, certification, and experience. The independent variable most closely related to student performance and growth on the pre and post writing assessments was the school where students were enrolled (either the MWTI program school where teachers were receiving professional development on the teaching of writing, or the comparison school where teachers were not receiving such professional development).

Dependent teacher variables included the degree of implementation of the strategies introduced in the professional development. Dependent student variables included the seven scores on pre and post writing assessments (six analytic and one holistic score), and the syntactic and prominent-feature measures.

Data Sources

Data sources are described in table 4 in relation to research questions and analysis.

Table 4. Research Questions and Related Data Sources, Measures, and Analysis

Research Questions	Data Sources	Measures	Analysis
Questions related to teacher practices			
To what extent, if at all, do teachers in their first year of MWTI professional development use the recommended strategies in their classrooms?	Semi-structured interviews in program and comparison schools. Semi-structured interviews with principals.	Degree of Implementation Continuum, which describes teaching practices along a continuum from 1 (most isolated and skill based) to 4 (most student centered and process oriented).	Qualitative analysis using descriptors on the Continuum of Implementation.
What differences, if any, exist in the teaching strategies of program versus comparison teachers?	Semi-structured interviews in program and comparison schools.	Numerical values derived from the Continuum of Implementation and the Classroom Strategy Measurement.	Analyses for differences between program and comparison teachers.
Questions related to student outcomes			
What differences, if any, exist in scores of program versus comparison students on pre and post writing assessments from a national scoring conference using the NWP Analytic Writing Continuum and holistic rubric?	Pre and post writing assessments from program and comparison students.	Scores from the national scoring conference using the NWP Analytic Writing Continuum and holistic rubric.	Statistical analysis for differences between and within groups.
What syntactic features are present in the pre and post writing assessments of upper elementary students.	Pre and post writing assessments from all program students and a random sample of comparison students.	Word count, T-unit count, dependent-clause count resulting in measures of mean T-unit length, and mean clause length.	Statistical analysis to determine differences within and between syntactic measures of both groups.
What prominent features are present in the pre and post writing assessments of upper elementary students?	Pre and post writing assessments of all program student writing and a random sample of comparison student writing.	Prominent-feature analysis.	Statistical analysis to determine differences in prominent features between pre and post writing assessments.

Questions related to correlations between measures		
What correlations, if any, exist between measures of syntactic maturity and score points on the NWP holistic rubric?	Syntactic measures. Scores based on the NWP holistic rubric.	Statistical analysis to determine whether or not correlations exist between syntactic measures and the holistic scores.
What correlations, if any, exist among the prominent features and the scores derived from the NWP Analytic Writing Continuum and holistic rubric?	Prominent-feature analysis. Scores based on the Analytic Writing Continuum and NWP holistic rubric.	Statistical analyses to determine correlations between prominent features and scores on the analytic and holistic rubrics.
What relationships, if any, exist among teacher practices and student outcomes?	Measures of teacher practices and student assessment outcomes.	Statistical analyses to determine Relationships between teacher practices and student outcomes.

Collection of Teacher and Student Data

The research team scheduled and conducted teacher interviews and classroom audits in April and May of the 2005–06 school year. The interview featured open-ended questions designed to encourage teachers to discuss their writing instruction. Suggestive words such as *revision* or *prewriting* were not used in the protocol. Instead, the questions asked for discussion: for example, “Tell me how you went about leading your students through [a particular assignment], from beginning to end.” A sample interview guide appears in Appendix A.

Student writing performance was assessed using a partially counterbalanced design through repeated measurement of on-demand writing exercises. Pre measures of writing were taken in September, with the program group being assessed three weeks later than the comparison group due to Hurricane Katrina. Post measures were taken one week apart, in late April (program group) and early May (comparison group). One writing prompt or task was selected from NWP’s archive of writing assessment tasks. (The NWP archive contains over 780 prompts, classified by a number of descriptive elements such as genre, modality, information source, degree of scaffolding, timing, and grade-level appropriateness. These elements were used to select the initial prompt for use in the research.) The second writing prompt was locally designed to be parallel to the NWP prompt in mode of writing, amount of scaffolding, and kinds of thinking required. The

two prompts were administered untimed by the classroom teachers following training by members of the research team. See Appendix B for writing prompts.

Data Analysis

Framework for Evaluating and Scoring Student Writing

To ensure technical rigor and credibility of the writing assessment data, scoring and data processing were conducted nationally, independently of the local sites. Building upon a long tradition of writing assessment, NWP provided a rigorous evaluation framework and conducted an independent national conference for scoring students' written work for Local Site Research Initiative studies nationwide.

Rubric

The evaluative framework employed an assessment system known as the NWP Analytic Writing Continuum (AWC), which was adapted from the *6+1 Traits of Writing* model (Culham, 2003). A national panel of experts on the assessment of student writing, along with NWP senior researchers, determined that the *6+1 Traits* framework, while sufficiently comprehensive, required certain modifications to make it more appropriate for use in the LSRI research studies. The resulting AWC is sufficiently unique to warrant its own nomenclature and identity. The following modifications were made in the rubric prior to the initial scoring conference in 2005:

- The scale of the rubric was extended from four to six points in order to ensure sufficient discrimination and therefore to allow increased sensitivity to any changes in student performance.
- The language defining the traits was clarified to enhance the reliability of evaluative judgments.
- The evaluative judgments were modified to focus exclusively upon the student's writing (where, on occasion, the original rubric included references to the reader's reactions or to the writer's personality as the basis for judgment).

A number of additional modifications were made in the assessment framework for use in 2006, refining and clarifying the definitions of the constructs measured. The revised *NWP Analytic Writing Continuum* rubric assessed the following elements of writing:

- ***Content*** (including quality and clarity of ideas and meaning)—The content category describes how effectively the writing establishes and maintains a focus; selects and integrates ideas related to content (i.e., information, events, emotions, opinions, and perspectives); and includes evidence, details, reasons, anecdotes, examples, descriptions, and characteristics to support, develop, and/or illustrate ideas.
- ***Structure***—The structure category describes how effectively the writing establishes logical arrangement, coherence, and unity within the elements of the work and throughout the work as a whole.
- ***Stance***—The stance category describes how effectively the writing communicates a perspective through an appropriate level of formality, elements of style, and tone appropriate for the audience and purpose.

- ***Sentence Fluency***—The sentence fluency category describes how effectively the sentences are crafted to serve the intent of the writing in terms of rhetorical purpose, rhythm, and flow.
- ***Diction (Language)***—The diction category describes the precision and appropriateness of the words and expressions for the writing task and how effectively they create imagery, provide mental pictures, or convey feelings and ideas.
- ***Conventions***—The conventions category describes how effectively the writing demonstrates age-appropriate comparison of usage, punctuation, spelling, capitalization, and paragraphing.

In addition to scores in each of these areas, each writing sample received an overall holistic score, one defined not as an aggregate of these component parts but as an independent, overall, summary judgment.

Scoring

Student writing samples from all four sites were integrated and scored at a national scoring conference—each local site research team determined which set of grade-level standards (elementary, middle, or high school) would be applied when scoring work from their site. Student writing for this study was scored using the elementary standards. All of the student writing was coded, with identifying information removed so that scorers could not know any specifics of the writing sample being evaluated (e.g., site of origin, group [program or comparison], or time of administration [pre or post]). Across all sites, 16% of the student writing was scored twice so that interrater reliabilities could be calculated.

The scorers participated in six hours of training using sets of exemplar papers that included “anchor papers”—samples exemplifying each level of achievement—along with descriptive commentary and practice papers for calibrating scorers throughout the scoring process. Overall, reliabilities (measured as interrater agreement, defining agreement as two scores that are identical or within one single score point of each other) ranged from 88 to 94%, with a median across all scores of 91%. The lowest reliability was observed for stance (88%) and the highest for the holistic scores (94%). This level of reliability is as high as is typically observed, and easily adequate to support the research purposes pursued here.

Syntactic and Prominent-Feature Analysis

In addition to the analytic and holistic scoring, all program school papers and a random sample of 10% of the comparison school papers were also analyzed by the MWTI research team for syntactic measures and prominent features. Those procedures are described in the following sections.

Selection of the Research Team

The members of the research team were selected by the principal investigator for 1) prior experience, 2) successful background in classroom teaching, 3) strong leadership, 4)

participation in and distinguished service to the National Writing Project. In all, the research team was composed of 5 individuals, including Dr. Swain and Dr. Graves.

Prominent-Feature Analysis

For the qualitative or prominent-feature analysis, team members were asked to report the prominent features of the writing. They were reminded to rely on their own professional experience to distinguish and identify prominent features. Every paper had at least one prominent feature; some had as many as ten. The process included partnered analysis during the early stages, with consensus for papers considered difficult. Fifteen percent of the papers were randomly selected for double reading.

During analysis it became apparent that every is either distinctly positive or distinctly negative. It also became apparent that certain features, such as repetition, may be either negative (as in redundancy, e.g., “I like I like....”) or positive (as in effective anaphora, e.g., Martin Luther King’s “I Have a Dream”). Moreover, some features may be negative (faulty spelling) without a positive corollary. Still others may be positive (the presence of a cumulative sentence) without a negative corollary. How this complexity was resolved is described below in the section Developing the Prominent-Feature Score. The complete list of prominent features identified in this study and their definitions is shown below. The features are organized by categories, and the negative features appear in italics.

SHOW WRITING

Elaborated details—use of vivid, appropriate, or striking details; goes beyond a listing of details

Sensory Language—language addressing the six senses, including direct quotations

Metaphors—all types of metaphoric language (metaphor, simile, personification); especially noted is the use of simple, common words used in metaphoric ways

VOCABULARY/DICTION

Vivid Verbs—striking diction, very appropriate and descriptive

Hyperbole—exaggeration

Striking Words—Striking word usage (other than verbs); appropriate or surprising adjectives, adverbs, etc.

Usage Problems—presence of social, racial, or geographical dialect; not “grammatical errors” or slips

Verb tense—inappropriate shift in tense or subject/verb disagreement

Inappropriate Adjectives—presence of inappropriate or overused adjectives

SENTENCE STRUCTURE

Cumulative Sentence—sentence with a base clause and one or more free modifiers

Verb Cluster—type of free modifier (-ing or -ed participle)

Absolute—type of free modifier; an independent noun with its own verb and deleted auxiliary verb

Adverbial Lead—beginning the sentence with striking adverbial (word, preposition, or clause)

Balance, Parallelism—all types of parallel structure

Repetition—repeating the same word, or a form of it, effectively (also includes repetition of phrases)

Striking Sentence—presence of an unusually mature sentence form

Variety—effective use of a variety of sentence forms and lengths

Dialogue—use of dialogue within a piece of writing

Weak Structural Core—sentences that are “derailed” with misplaced awkward elements.

Also includes sentence fragments. “...putting the heart of [the] idea into prepositional phrases, object noun clauses, adjectives, adverbs or other ancillary parts of the sentence, wasting the subject and/or verb position on indefinite, evasive expressions.... [the] habitual wasting of the subject-verb position....” (Krishna 1975).

Garble—unintelligible sentences

Choppy—pattern of short, often redundant, sentences

ORGANIZATION

Well-organized—clear pattern of organization

Chronological Sequence—organizational structure based on events in time order

Transition Words—the presence of key function words to enhance organization. (See Becker 1965.)

Supporting Ideas—inclusion of multiple ideas to support a theme

Chain of Ideas—ideas presented so that one leads to the next

Central Idea—a dominant theme around which the writing is organized

Supporting Details—citing of details to support a point

Beginning—inclusion of an introduction

Ending—a conclusion or concluding statement

Poorly Organized—obvious lack of organization

Redundant—repeating the same idea or concept over and over; sometimes described as “verbage” or “mindless filler”

List Technique—simple list of words or phrases; undeveloped writing

Unfocused—a paper with no apparent central idea

Underdeveloped—a paper that presents an idea, but does not develop it

OTHER

Voice—the presence of an original, personal, or authentic conception of the subject

Onomatopoeia—the use of words to replicate sounds

Rhythm and Flow—the arrangement of words such that a rhythm or parallelism is established

Aside to Reader—the use of parentheticals to engage the reader (e.g., of course, as I said)

Faulty Spelling—a pattern of faulty spelling

Poor Punctuation—a pattern of faulty or missing age-appropriate punctuation

Illegible Handwriting—handwriting so faulty it is unreadable, or almost unreadable

Addresses Reader—Inappropriately addressing the reader (I am going to tell you . . .)

Developing the Prominent-Feature Score

The analysis conducted by the research team revealed the following:

1. The prominent features of a piece may be either positive or negative. The research team settled on 28 positive features and 15 negative features, creating a possible range of 43 score points.
2. Each prominent feature was assigned the value of +1 or -1, indicating positive or negative. No feature was rated more or less valuable than another.
3. Ten points were added to each paper's score. This resulted in all prominent-feature scores being reported as positive numbers while maintaining the full scoring range of 43 points.

This possible range of 43 points is far greater than the range provided in conventional writing assessment, and thus allows for more nuanced assessment interpretation of each piece of student writing. Furthermore, beyond its numerical value each paper has a set of prominent features, verbal descriptors which themselves can serve as a valuable diagnostic tool for individual students' growth in writing. An added value of this process for staff developers is the ability to assess the success of teachers' prior professional development programs, and to determine teachers' needs for future programs.

Syntactic Measures Analysis

For the quantitative syntactic measures analysis (Hunt 1963), team members counted the number of words, number of dependent clauses, and number of T-units (independent clauses plus related dependent clauses). When problematic passages occurred, team members asked the leaders for assistance. Papers were analyzed twice, especially for the accurate identification of dependent clauses.

Measures of Syntactic Maturity

Syntactic measures were derived from the number of words, number of T-units, mean T-unit length, number of dependent clauses, and mean clause length. Mean T-unit length was derived by the formula:

$$\frac{\text{Number of Words}}{\text{Number of T-units}}$$

Mean clause length was derived by the formula:

$$\frac{\text{Number of Words}}{\text{Number of T-units} + \text{Number of Dependent Clauses}} \text{ (Hunt, 1965)}$$

Analysis of NWP Strategies in Classroom Practice: The Sum of Strategies Score

Members of the research team double-scored each of the teacher interviews for evidence of ten NWP teaching strategies. These included: 1) student choice in the selection of topics for writing; 2) emphasis on reading-writing connections; 3) time devoted to prewriting activities or brainstorming activity; 4) opportunities to read one's writing aloud to peers; 5) teacher-student conferences to discuss writing in process; 6) mini-lessons to teach directly to specific student need and interest; 7) encouragement to revise or try out several approaches to a certain topic; 8) close editing for diction, mechanics, and syntax; 9) publishing student writing in various outlets and venues beyond the classroom; and 10) modeling examples of good writing.

Three codes were used to indicate whether or not the ten teaching strategies were being implemented in each classroom: a zero (0) for no mention of a strategy, a one (1) for mention but no evidence of the strategy being used in the classroom, and a two (2) for strong evidence of that strategy at play in the classroom. Discrepancies among these scores were ameliorated through consensus. Each teacher received a summary score, ranging from 0 to 20, indicating the level of inclusion of NWP strategies in their practice. This score is called the sum of strategies score. Definition of strategies appears in appendix C.

Analysis of Teachers' Degree of Implementation

To describe variations in the degree to which individual teachers implemented strategies advocated in the professional development, the research team used a degree of implementation continuum (see Appendix D). The continuum is a four-point scale reflecting a continuum of classroom practice, from teacher-centered/skill-based at one end of the scale, to student-centered/reflective practice at the other. This continuum takes into account data from interviews and classroom audits or observations. A score of 4 represents full implementation and practice of NWP process strategies; a score of 1 represents a traditional approach to the teaching of writing. The resulting score is called the *degree of implementation*. Each interview was scored two times independently by members of the team. Because the scores represent a continuum of practice, adjacent scores were considered to be in agreement and were averaged to obtain a final score. Interviews that received scores that differed by two points were to be rescored by the principal investigator; however, there were no discrepancies within the set of 19 interviews for the present study, thus interrater reliability for degree of implementation was 1.0.

Results and Findings

The Effect of Professional Development on Teacher Practices

This section describes how writing is taught in the program and comparison schools, as the differences in students' writing seen between the two groups may be directly attributed to the difference in the way writing is taught. Eleven of the 26 teachers in the program group and 8 of the 16 in the comparison group were selected for this part of the

study (a total of 19 teachers). Each of these 19 teachers participated in an intensive one-hour interview. Both principals were also interviewed to corroborate interview results.

Use of Strategies

Table 5 details practices from the classrooms of program and comparison teachers who were interviewed. This table allows researchers to consider which of the practices presented in the professional development were most and least readily adopted by teachers in the professional development. It has implications for professional developers in planning future programs. Prewriting and editing were most readily implemented by both program and comparison teachers, although program teachers described more complete implementation. Minilessons were least readily implemented among the program teachers, although they used the mini-lesson strategy to a greater extent than the comparison teachers. Modeling, a strategy to help students understand what is expected in a writing assignment, was not implemented at all among comparison teachers.

Table 5. Level of Implementation, Process Teaching Strategies, Program and Comparison Groups

Category of NWP Writing Process Curriculum	Program Group <i>n</i> = 11				Comparison Group <i>n</i> = 8			
	Frequency			Mean Score	Frequency			Mean Score
Implementation of Strategies 0 = no implementation; 1 = some implementation; 2 = full implementation	0	1	2		0	1	2	
Student Choice	4	1	6	1.18	4	3	1	.63
Reading/Writing Connections	4	4	3	.91	5	3	0	.39
Prewriting	0	2	9	1.82	3	2	3	1.0
Peer Response	2	4	5	1.27	3	4	1	.75
Teacher Conference	1	3	7	1.54	4	3	1	.63
Mini-Lessons	4	6	1	.73	7	1	0	.13
Revision	1	6	4	1.27	6	2	0	.25
Editing	1	2	8	1.64	1	5	2	1.13
Publishing	0	9	2	1.18	2	4	2	1.0
Modeling	2	4	5	1.27	8	0	0	0
Overall Means for Program and Comparison				1.3				.6

Note: *N* = 19.

The overall means of implementation of all strategies among the two groups indicate that the program group benefited from the professional development with a mean of 1.3 out of a possible 2 as opposed to the comparison group's mean of 0.6. Specific differences in approaches to the teaching of writing between the program and comparison groups, based on the 10-category classification of the NWP writing process model, are further described below.

1. **Student Choice of Topics for Writing.** The program teachers offered more opportunities for student choice for writing topics than did the comparison teachers. Table 5 shows the difference of .55 on a 2-point scale. Interestingly four program teachers and four comparison teachers did not allow for student choice. Six of the eleven program teachers fully implemented student choice.
2. **Reading-Writing Connections.** Program teachers made more connections between reading and writing than the comparison teachers (.91 for the program, .39 for comparison) on the 2-point scale. Program teachers were fairly equally distributed among the three categories: three fully implemented the connections, four somewhat implemented the connections, and four made no connections. In the comparison group, five of the eight teachers indicated no connections, three mentioned the connections, but evidence of implementation was lacking.
3. **Prewriting.** All the program teachers provided some prewriting activities for their students: nine fully implementing and two partially implementing. Three of the eight comparison teachers fully implemented and two partially implemented prewriting. Program teachers' mean score was 1.82; comparison teachers, 1.0.
4. **Peer Response.** Peer response provides the writer the opportunity to share writing in process and then receive feedback from peers. Program teachers scored 1.27 out of a possible 2.0 on implementation of this strategy; comparison teachers, .75.
5. **Teacher Conferences.** Ideally teacher conferences serve more than a remedial function, providing all students with one-on-one instruction. With 8 of the 11 teachers in the program fully implementing conferences, their mean was 1.54, compared to .63 for comparison teachers.
6. **Minilessons.** These brief, focused lessons are designed to meet student needs and to interest them in writing. Program teachers' mean for implementation was .73 for minilessons, their lowest score. Only one comparison teacher fully implemented this approach, six partially implemented it, and four had no implementation, for a mean of .13.
7. **Revision.** Ten of the eleven program teachers implemented a revision process in which students reexamined their conception and presentation of ideas. Among comparison teachers, six indicated no attention to revision, and two indicated only partial implementation. The mean for program teachers was 1.27 on the 2.0 scale, and for comparison teachers, .25.
8. **Editing.** Although teachers in both groups indicated emphasis on careful reading with attention to diction, mechanics, and syntax, program teachers' mean was higher than the comparison teachers': 1.64 versus 1.13.
9. **Publishing.** This involves publishing student writing outside the classroom. In this category, the difference between the two groups was the smallest of all the differences: 1.18 for program teachers, 1.0 for comparison teachers.
10. **Modeling.** This refers to the teacher modeling various aspects of writing for students. Here the program group implementation is clearly superior, five teachers fully implemented the approach, four partially implemented it, and two indicated no implementation for a mean of 1.27. In the comparison school, none of the eight teachers implemented modeling at all.

Sum of Strategies Score

To obtain this score for each teacher, we add individual teachers' scores on implementation of the ten strategies. The actual range of scores was 1–19 on a 20-point scale, with program teachers' scores ranging from a high of 19 to a low of 6. Comparison teacher scores ranged from a high of 11 to a low of 1. Means for the two groups indicate that program teachers, with a mean of 12.7, implemented the strategies more fully than comparison teachers, whose mean was 5.9. Table 6 shows the frequency of the sum-of-strategies scores of the teachers interviewed.

Table 6. Sum of Strategies Score

Score	Program Frequency <i>n</i> = 11	Comparison Frequency <i>n</i> = 8
18–20	2	0
15–17	2	0
12–14	3	0
9–11	2	3
6–8	2	1
3–5	0	2
0–2	0	2
Mean	12.7	5.9

Degree of Implementation

The strategy scores indicate which of the writing project teaching strategies were practiced in the classrooms. The research team also analyzed the teacher interviews to determine a degree of implementation score for each individual teacher. The interview guide was designed to gather information about the extent to which teachers practiced process-based student-centered instructional strategies. In full implementation these strategies include teaching based on student need, teaching skills in the context of authentic student writing, the teacher writing and modeling his or her own writing process, effective use of peer response to writing, reflective practice, and opportunities for publishing outside the school.

The interview guide was also designed to gather information on what may be classified as a traditional or standard approach to writing instruction. Reduced to its simplest form, this approach includes the following: 1) topics for writing are assigned by teachers (students often all write to the same topic); 2) students' papers are turned in to the teacher at an assigned time; 3) the papers are read by the teacher, who assigns each a grade; and 4) the papers are returned to the students.

Each of the 19 teachers received a score of 1 to 4, with 4 being highest, based on the levels described by the continuum of implementation (see Appendix D). At Level 4 the continuum identifies those teachers who employ many if not all of the characteristics of a process curriculum, and at Level 1 those who employ a more standard or traditional approach. Levels 2 and 3 represent gradations between these two approaches, thus the

rationale for identifying the instrument as a “continuum.” Frequency of the degree of implementation scores for the teachers in the program and comparison groups is shown in table 7.

Table 7. Frequency of Teachers’ Degree of Implementation Scores for Program and Comparison Teachers

Degree of Implementation	Number Comparison Teachers	Number Program Teachers
Score	<i>n</i> = 8	<i>n</i> = 11
1	5	1
2	1	3
3	2	5
4	0	2
Mean	1.6	2.7

N = 19

As shown in table 7, the highest mean score was achieved by the program teachers who had participated in the professional development. Since the 2005–2006 school year was the first experience with writing project teaching strategies for the program teachers, their degree of implementation overall was considered satisfactory, with 7 (64%) of the program teachers scoring at Level 3 or higher, while 5 (63%) of the comparison teachers scored at Level 1. The scores indicate that MWTI professional development had a positive effect on the teaching practices of the program teachers.

Student Outcomes

This section describes student outcomes on four measures. Analytic and holistic writing scores are given for both pre and post writing assessments for all matched pairs in both the program and comparison groups. Prominent-feature analysis and measures of syntactic maturity are given for all matched pairs of program students and for a random sample of 10% of the comparison students.

Student Writing Performance, Holistic and Analytic

Papers from each occasion, pre and post writing assessment, were scored analytically and holistically, yielding a total of seven scores per occasion, each on a scale from 1 to 6. The areas receiving analytic scores were *content*, *structure*, *stance*, *sentence fluency*, *diction*, and *conventions*. The seventh score was the holistic rating.

Partial or complete data sets were available from 908 students: 565 at the program school and 343 at the comparison school. In the program school, the counts by grade were 201, 172, and 192 (grades 3–5, respectively). In the comparison school, the counts for the three grades were 116, 108, and 119 (grades 3–5 respectively). However, only 435 program participants and 217 comparison participants had both complete pre and post scores available. It is on these 652 participants that the following analyses are based.

The two groups had very similar scores on the pre writing assessment, with the comparison group scoring higher on five of the seven measures; summary statistics by group are given in table 8. Initial multivariate tests of pre/post differences were conducted separately for the program and comparison groups. Overall, program students showed statistically significant improvement in the set of scores (and on each score individually) from pre to post, whereas the comparison students' scores were essentially unchanged and were statistically indistinguishable across occasions. The multivariate test comparing pre to post scores for program students yielded $F(7, 428) = 19.15, p < .001$; for comparison students the result was $F(7, 210) = 1.52, p = .161$. Table 9 summarizes the statistical comparisons from pre to post writing assessment by group for each of the scores.

Consistently, the largest gains from pre to post writing assessment were observed for program participants in the third grade, averaging over a full point per score overall. Gains were noted for program students in the other grade levels as well, though not as large. Fourth grade students' gains averaged about one-half point; fifth graders' gains averaged about two-tenths of a point. Table 10 contains pre and post writing assessment scores by grade level for the program participants. Among the comparison site participants, a consistent gain was observed for the third grade students, averaging about three-tenths of a point. For comparison students, the pre to post writing assessment changes were essentially nil for fourth grade and declined by about one-quarter point on average in the fifth grade. Table 11 includes the individual scores by occasion and by grade for the comparison students.

The estimated effect sizes, shown in table 9, indicate that program students' scores for each of the analytic traits improved on average by at least one-half a standard deviation to over six-tenths of a standard deviation, while the holistic score improved by nearly six-tenths of a standard deviation. These differences are both statistically dependable and large enough to warrant attention. The comparison students, who presumably did not receive similar writing instruction and practice opportunities, showed changes from pre to post writing assessment that were close to nil. All of the comparison group effect size estimates are small in absolute value; none of their pre to post differences was statistically different from zero.

Table 8. Summary Statistics for Scores by Group

Score	Program: Pre assessment	Program: Post assessment	Program Difference (Post – pre)	Comparison: Pre assessment	Comparison: Post assessment	Comparison Difference (Post – pre)
Content	2.6 (1.1)	3.2 (1.1)	.6	2.6 (1.0)	2.6 (1.0)	0
Structure	2.3 (1.0)	2.9 (1.2)	.6	2.4 (1.0)	2.4 (1.0)	0
Stance	2.6 (1.2)	3.3 (1.2)	.7	2.6 (1.1)	2.6 (1.1)	0
Sentence fluency	2.5 (1.1)	3.1 (1.2)	.6	2.6 (1.0)	2.6 (1.0)	0
Diction	2.5 (1.1)	3.2 (1.1)	.7	2.5 (1.0)	2.6 (1.0)	.1
Conventions	2.4 (1.0)	3.0 (1.1)	.6	2.5 (1.0)	2.6 (1.0)	.1
Holistic	2.5 (1.1)	3.2 (1.1)	.7	2.6 (1.0)	2.6 (1.0)	0

Notes: Mean values are given; values in parentheses are standard deviations. $N = 435$ for program, 217 for comparison group.

Table 9. Statistical Comparisons of Overall Performance by Group

Score	Program: 95% CI for pre post difference, <i>ES</i>	Comparison: 95% CI for pre post difference, <i>ES</i>
Content	(0.48, 0.74) 0.56	(-0.23, 0.10) -0.07
Structure	(0.48, 0.74) 0.60	(-0.19, 0.16) -0.02
Stance	(0.50, 0.78) 0.55	(-0.16, 0.18) 0.01
Sentence fluency	(0.47, 0.73) 0.53	(-0.12, 0.22) 0.04
Diction	(0.56, 0.81) 0.64	(-0.08, 0.23) 0.08
Conventions	(0.48, 0.70) 0.57	(-0.04, 0.26) 0.11
Holistic	(0.51, 0.77) 0.58	(-0.16, 0.16) 0.00

Notes: Confidence intervals (*CI*s) and effect size (*ES*) estimates are constructed so that positive values reflect higher scores on post writing assessments than on pre assessments. If lower and upper limits of *CI* have the same sign and do not include zero, then the difference between pre and post writing assessment scores means is statistically different from zero at the .05 level. *ES* estimates are post assessment mean minus pre assessment mean divided by pre assessment *SD*. *N* = 435 for program, 217 for comparison group.

Table 10. Summary Statistics for Scores by Grade for Program Participants

Score	Pre Assessment (Grade 3)	Post Assessment (Grade 3)	Pre Assessment (Grade 4)	Post Assessment (Grade 4)	Pre Assessment (Grade 5)	Post Assessment (Grade 5)
Content	2.2 (0.9)	3.2 (1.2)	2.8 (1.2)	3.3 (1.2)	2.9 (1.0)	3.2 (0.9)
Structure	1.9 (0.8)	2.9 (1.3)	2.5 (1.1)	3.0 (1.2)	2.8 (0.9)	3.0 (0.9)
Stance	2.1 (1.0)	3.2 (1.2)	2.8 (1.2)	3.3 (1.3)	3.1 (1.1)	3.3 (0.9)
Sentence fluency	1.9 (0.9)	3.0 (1.3)	2.7 (1.2)	3.1 (1.3)	3.0 (1.0)	3.2 (0.9)
Diction	2.0 (0.9)	3.2 (1.2)	2.7 (1.1)	3.2 (1.3)	3.0 (0.9)	3.3 (0.9)
Conventions	1.9 (0.9)	3.0 (1.2)	2.6 (1.1)	3.0 (1.2)	2.8 (0.8)	3.1 (1.0)
Holistic	2.0 (0.9)	3.1 (1.2)	2.8 (1.2)	3.3 (1.2)	3.0 (1.0)	3.1 (0.9)

Notes: Mean values are given; values in parentheses are standard deviations. *N* = 161 for grade 3, 145 for grade 4, and 129 for grade 5.

Table 11. Summary Statistics for Scores by Grade for Comparison Participants

Score	Pre Assessment (Grade 3)	Post Assessment (Grade 3)	Pre Assessment (Grade 4)	Post Assessment (Grade 4)	Pre Assessment (Grade 5)	Post Assessment (Grade 5)
Content	2.1 (0.9)	2.3 (1.0)	3.0 (0.9)	2.9 (1.1)	2.9 (1.0)	2.6 (0.9)
Structure	1.9 (0.8)	2.1 (0.9)	2.7 (1.0)	2.7 (1.1)	2.8 (0.9)	2.4 (1.0)
Stance	2.0 (0.9)	2.3 (1.1)	3.0 (1.0)	3.0 (1.2)	3.0 (1.0)	2.7 (1.0)
Sentence fluency	2.0 (0.9)	2.4 (0.9)	2.8 (0.9)	2.8 (1.1)	3.0 (1.0)	2.7 (1.0)
Diction	1.9 (0.9)	2.2 (0.9)	2.8 (0.9)	2.8 (1.1)	2.9 (0.8)	2.7 (0.9)
Conventions	1.9 (0.9)	2.3 (1.0)	2.8 (0.9)	2.8 (1.1)	2.8 (0.8)	2.7 (0.8)
Holistic	2.0 (0.8)	2.3 (0.9)	2.9 (0.9)	2.9 (1.0)	2.9 (1.0)	2.6 (0.9)

Notes: Mean values are given; values in parentheses are standard deviations. *N* = 75 for grade 3, 73 for grade 4, and 69 for grade 5.

Summary of One-Between, One-Within ANOVA Analyses

The following tables summarize one-between, one-within subjects factors ANOVA analyses of the writing scores from the 2006 study. The between-subjects factor was program condition (program school or comparison school) and the within-subjects factor was assessment occasion (pre or post). Interpretation of the main effect of program in this type of design is not informative unless the groups are equivalent on the pre assessment occasion, since it combines both pre and post assessment scores for each program group.

Likewise, the main effect of occasion lumps both program and comparison groups together and therefore may not indicate what happened to either group separately. The test of interaction (group by occasion) conveys whether there is differential change from pre to post writing assessment by group; further testing (e.g., via simple effects tests) would be required in order to assert whether pre/post changes were significantly different from zero for individual groups.

Earlier analyses suggested that the groups were roughly comparable on pre assessment scores and that, with the exception of third grade, only the program students showed statistically dependable gain from pre to post writing assessment. Therefore, the main effect of group comparisons and the group-by-occasion interactions are essentially telling the same story here—that the difference between groups is principally due to the fact that only the program students showed a change in performance, improving from pre to post writing assessment, whereas the comparison students showed no consistent change. As discussed, the third grade comparison students did improve, though not nearly to the degree that their program counterparts did.

Program students showed statistically significant improvement in the overall set of scores (and on each score individually) from pre to post writing assessments in relation to the comparison students' scores, which were essentially unchanged and were statistically indistinguishable across occasions. Table 12 summarizes the results of a repeated measures ANOVA of the pre and post writing assessments for program and comparison groups for each attribute of writing as well as for the holistic assessment.

For each set of scores, there was a significant difference at the .001 level for occasion, interaction, and six of the seven measures for group. The other measure of significance for group was $p = .008$ for conventions. There was also a significant difference in program students' own scores between pre and post assessments. The significant difference in the interaction between the occasion (pre or post) and the group (program or comparison) indicates that the difference is due to group. Table 12 indicates that the significant differences in all areas of writing that were assessed were due to the program. In brief, growth in all areas of writing was significantly greater for the program group between the pre and post writing assessments, *and* significantly greater than that of the comparison group.

Student Writing Performance, Syntactic Measures, and Prominent Features ***Frequency of T-units, Clauses, and Prominent Features***

One or both prominent-feature scores were available for most of the program students, though only for a sample of the comparison students. Subsequent analyses are reported only for those students having both pre test and post test prominent-feature scores. There were 467 program students and 37 comparison students with both sets of scores ($N = 504$). Because there were so few comparison students with prominent-feature scores, any conclusions about program students versus comparison students must be considered tentative.

Table 12. Repeated-Measures ANOVA Results for All Matched Cases on Holistic and Analytic Scores

Score	Variance Component	df	Mean Square	F Ratio	Test of Significance <i>P</i> (<i>F</i>)	Effect Size
Holistic	Between subjects Program group (pre/post)	1	19.857	13.742	<.001	.021
	Error (between)	650	1.445			
	Within subjects Occasion (pre, post)	1	29.565	33.053	<.001	.048
	Group x Occasion	1	30.565	33.053	<.001	.048
	Error (within)	650	0.894			
Content	Between subjects Program group (pre/post)	1	22.822	15.660	<.001	.024
	Error (between)	650	1.457			
	Within subjects Occasion (pre, post)	1	21.205	24.358	<.001	.036
	Group x Occasion	1	32.969	37.872	<.001	.055
	Error (within)	650	0.871			
Structure	Between subjects Program group (pre/post)	1	15.794	11.369	<.001	.017
	Error (between)	650	1.389			
	Within subjects Occasion (pre, post)	1	25.659	28.291	<.001	.042
	Group x Occasion	1	28.515	31.440	<.001	.046
	Error (within)	650	0.907			
Stance	Between subjects Program group (pre/post)	1	25.804	16.029	<.001	.024
	Error (between)	650	1.610			
	Within subjects Occasion (pre, post)	1	31.294	31.458	<.001	.046
	Group x Occasion	1	28.718	28.868	<.001	.043
	Error (within)	650	0.995			
Sentence Fluency	Between subjects Program group (pre/post)	1	8.986	5.571	<.001	.008
	Error (between)	650	1.613			
	Within subjects Occasion (pre, post)	1	30.109	33.706	<.001	.049
	Group x Occasion	1	22.119	24.761	.,001	.037
	Error (within)	650	0.893			
Diction	Between subjects Program group (pre/post)	1	31.824	22.004	<.001	.033
	Error (between)	650	1.446			
	Within subjects Occasion (pre, post)	1	41.806	50.402	<.001	.072
	Group x Occasion	1	26.748	32.248	<.001	.047
	Error (within)	650	0.829			
Conventions	Between subjects Program group (pre/post)	1	9.907	6.392	.008	.010
	Error (between)	650	1.550			
	Within subjects Occasion (pre, post)	1	35.146	51.731	<.001	.074
	Group x Occasion	1	17.015	25.043	<.001	.037
	Error (within)	650	0.679			

Note: *ES* is partial eta-squared. Program $n = 435$; comparison $n = 217$.

Summary statistics for the writing samples are given for the program school students by grade in table 13. Similar information for the comparison students is given in table 14. For program students ($n = 467$), there was a statistically dependable increase in the set of four summary indicators (number of words, number of T-units, number of clauses, and prominent-feature scores) from pre to post writing assessment, multivariate $F(4, 463) = 15.090, p < .001$. Individually, the increases in number of words ($M = 37.9, SD = 129.7$), number of T-units ($M = 3.0, SD = 15.4$), number of clauses ($M = 1.7, SD = 6.7$), and prominent-feature score ($M = 0.5, SD = 4.5$) were also statistically significantly different from zero ($p < .001$ for all except prominent features, $p = .021$). Thus we may conclude that program school students wrote longer, more complex responses on the post assessments than on the pre assessment. The same could not be said for the comparison students. The multivariate comparison of pre and post summary scores showed no statistically significant difference, $F(4, 33) = 2.137, p = .098$. With the exception of the prominent-feature total score, each of the four indicator means was lower on the post assessment than on the pre assessment (see table 14). Whether this suggests a lack of motivation or some other reason unrelated to writing skill is unclear. What is clear is that the program students, though comparable to the comparison students at pretest, generated longer essays on the post writing assessment.

Table 13. Summary Statistics for Program School Writing Samples by Grade

Grade	<i>Pre Assessment</i>			PF Score	<i>Post Assessment</i>			PF Score
	Words	T-units	Clauses		Words	T-units	Clauses	
3	96 (75)	12 (9)	3 (4)	10 (3)	156 (100)	18 (12)	5 (4)	11 (3)
4	137 (78)	15 (9)	5 (4)	11 (3)	188 (174)	21 (21)	7 (7)	11 (4)
5	174 (90)	18 (10)	8 (5)	13 (3)	178 (108)	16 (10)	8 (7)	13 (4)
All	135.5	15.4	5.1	11.2	173.4	18.4	6.8	11.7

Notes: Means and SDs (in parentheses) are rounded to nearest integer. Results are based on $n = 163, 145,$ and 159 students in grades 3, 4, and 5, respectively.

Table 14. Summary Statistics for Comparison School Writing Samples by Grade

Grade	<i>Pre Assessment</i>			PF Score	<i>Post Assessment</i>			PF Score
	Words	T-units	Clauses		Words	T-units	Clauses	
3	104 (98)	12 (9)	4 (4)	9 (2)	69 (27)	9 (4)	3 (2)	9 (2)
4	113 (52)	13 (6)	5 (4)	11 (2)	86 (33)	15 (6)	6 (7)	10 (2)
5	177 (99)	18 (11)	8 (5)	10 (3)	146 (71)	15 (6)	6 (7)	13 (3)
All	136.9	14.7	5.8	10.5	106.0	11.5	4.9	10.7

Notes: Means and SDs (in parentheses) are rounded to nearest integer. Results are based on $n = 9, 13,$ and 15 students in grades 3, 4, and 5, respectively.

Mean T-Unit Length Comparisons

Table 15 contains summary statistics by program condition and by grade for the mean length (in words) of T-units at pre assessment and post assessment. Significance tests, comparing pre to post writing assessment results, are given in table 16. In the third grade, there was virtually no difference in pre assessment scores across groups, but the program group students showed a statistically significant increase from pre to post writing assessment (see table 16). For fourth grade participants, only the comparison group students demonstrated a statistically significant increase in mean T-unit length from pre to post writing assessment. Fifth grade program participants showed a statistically significant increase in mean T-unit length, whereas there was no statistically dependable change different from zero for the comparison participants.

Program participants therefore had statistically longer average T-units, measured in number of words, at post assessment than at pre assessment in grades 3 and 5, and overall (see table 16). The effect sizes, expressed as standardized mean differences, were small (0.25) for grade 3, moderate to large (0.74) for grade 5 and negligible for grade 4 (0.08). Comparison participants had a moderate-to-large improvement in mean T-unit length for grade 4 ($ES = 0.71$), a small-to-medium decline in grade 3 (-0.36), and a small-to-medium increase in grade 5 (0.31). The very low number of cases having both pre and post writing assessment data for the comparison group means we should treat these latter statistics as tentative findings.

Table 15. Summary Statistics for Mean T-unit Length in Words by Grade and Program

Group	Pre test (Grade 3)	Post test (Grade 3)	Pre test (Grade 4)	Post test (Grade 4)	Pre test (Grade 5)	Post test (Grade 5)
Program	8.1 (2.8)	8.8 (2.7)	9.3 (2.6)	9.5 (2.1)	9.7 (2.4)	11.5 (2.9)
Comparison	8.1 (2.1)	7.4 (1.5)	8.7 (1.9)	10.0 (2.4)	10.3 (1.7)	9.8 (2.1)

Notes: Mean values are given; values in parentheses are standard deviations. Program $n = 163$ for grade 3, 145 for grade 4, and 159 for grade 5. Comparison $n = 9, 13,$ and 15 for grades 3–5, respectively.

Table 16. Statistical Comparisons of Mean T-unit Length by Group

Grade (n)	Program: 95% CI for pre test–post test difference, ES	Comparison: 95% CI for pre test–post test difference, ES
3 (163, 9)	(0.15, 1.24) * 0.25	(-2.51, 0.99) -0.36
4 (145, 13)	(-0.30, 0.70) 0.08	(0.33, 2.34) * 0.71
5 (159, 15)	(1.26, 2.33) * 0.75	(-2.14, 1.06) 0.31
Combined	(0.61, 1.22) * 0.34	(-0.77, 0.90) 0.03

Notes: Confidence intervals (CIs) and effect size (ES) estimates are constructed so that positive values reflect higher scores on post assessment than on pre assessment. * If lower and upper limits of CI have the same sign and do not include zero, then the difference between pre assessment and post assessment means is statistically different from zero at the .05 level. ES estimates are post assessment mean minus pre assessment mean divided by pre test SD . Sample sizes (ns) are given in parentheses by grade.

Table 17 gives a similar analysis using the one-between (program condition), one-within (occasion: pre test, post test) comparison on the combined grades data. Overall, we see no difference due to group, occasion, or group by occasion. This is due partly to (a) the relatively small sample size for comparison participants and (b) the inconsistent results across grades, as outlined in tables 15 and 16 and explained above.

Table 17. Repeated-Measures ANOVA Results for all Matched Cases on Mean T-unit Length

Source	<i>df</i>	<i>MS</i>	<i>F</i>	prob(> <i>F</i>)	<i>ES</i>
Between subjects					
Program group	1	4.893	0.524	.470	.001
Error (between)	502	9.342			
Within subjects					
Occasion (pre, post)	1	16.428	2.935	.087	.006
Group x Occasion	1	12.447	2.224	.137	.004
Error (within)	502	5.597			

Note: *ES* is partial eta-squared. Program $n = 467$; comparison $n = 37$.

Correlation of Prominent-Feature Scores and Length Indicators

How well the prominent-feature scores correlated with the length indicators (number of words, number of T-units, number of clauses) and NWP holistic scores is summarized by occasion in table 18. These results are based on the combined program and comparison groups having both pre test and post test scores for prominent features and the NWP analytic scores. One important conclusion that may be drawn from this table is that the correlation of the prominent-feature score was stronger with paper-length indicators (words, T-units, clauses) and the NWP holistic score on the pre test than on the post test. This could be explained in part by the changes in length indicators by group: post test essays were generally longer than pre test essays for program students, but were generally shorter for comparison students.

Table 18 . Correlations of Prominent-Feature Scores with Other Measures by Occasion

<i>Measure</i>	(1)	(2)	(3)	(4)	(5)
Prominent feature (1)		.24	.17	.25	.46
Number of words (2)	.53		.94	.76	.46
T-unit (3)	.49	.91		.59	.40
Number of clauses (4)	.39	.69	.53		.40
NWP holistic score (5)	.58	.59	.54	.43	

Note: Values in lower triangle are pre assessment correlations; those in upper triangle are post assessment correlations.

Pre and Post Frequencies of Prominent Features

Table 19 summarizes the relative frequency of appearance for each of the prominent-feature elements that was recorded; these figures represent those students ($N = 504$) having both pre and post test prominent-feature scores available (both treatment and comparison students). Among the more noteworthy changes across occasions were increases in some of the positive elements, such as use of sensory language, adverbial leads, use of supporting details, and, most sharply different of all, evidence of a central idea (from 0 to 33%). Among the negative elements, the sharpest decline was observed for the incidence of poor punctuation (from 20% to 10%). However, several positive features, for whatever reason, were observed to decline in frequency from pre test to post test. These included beginning (from 51% to 42%) and inclusion of supporting ideas (from 40% to 10%, though with a concomitant increase in use of supporting details). Most of the other elements remained roughly comparable across occasions in their frequency.

By frequency, several of the prominent features were rarely in evidence among the student essays (under 5% of the time). These include the positive features of vivid verbs, hyperbole, absolutes, transition words, chain of ideas, and onomatopoeia; several negative features that were also sparse on the landscape included change in verb tense, garbles, illegible handwriting, and inappropriate adjectives.

Table 19. Frequencies of Prominent-Feature Elements in Student Papers by Occasion for Combined Sites

Element	Attribute Type	Pre Test Frequency	Post Test Frequency
Elaborated details	Positive	70 (14%)	64 (13%)
Sensory language	Positive	13 (3%)	57 (11%)
Metaphors	Positive	28 (6%)	39 (8%)
Vivid verbs	Positive	18 (4%)	10 (2%)
Hyperbole	Positive	0	8 (2%)
Striking words	Positive	40 (8%)	35 (7%)
Change in verb tense	Negative	9 (2%)	10 (2%)
Usage problems	Negative	33 (6%)	18 (4%)
Cumulative sentences	Positive	8 (2%)	29 (6%)
Verb clusters	Positive	3 (1%)	23 (5%)
Dialogue	Positive	23 (5%)	48 (10%)
Absolutes	Positive	2 (0%)	4 (1%)
Adverbial leads	Positive	32 (6%)	62 (12%)
Balance/parallelism	Positive	23 (5%)	39 (8%)
Repetition	Positive	32 (6%)	18 (4%)
Striking sentence	Positive	49 (10%)	51 (10%)
Sentence variety	Positive	71 (14%)	48 (10%)
Beginning	Positive	257 (51%)	209 (42%)
End	Positive	168 (33%)	152 (30%)
Weak structural core	Negative	50 (10%)	37 (7%)
Choppy	Negative	47 (9%)	62 (12%)
Garbles	Negative	3 (1%)	1 (0%)
Well-organized	Positive	48 (10%)	46 (9%)
Chronological organization	Positive	59 (12%)	44 (9%)
Transition words	Positive	22 (4%)	18 (4%)
Supporting ideas	Positive	204 (40%)	48 (10%)
Supporting details	Positive	122 (24%)	204 (40%)
List of details/ideas	Negative	130 (26%)	162 (32%)
Weak organization	Negative	38 (8%)	44 (9%)
Redundant	Negative	168 (33%)	125 (25%)
Under-developed	Negative	186 (37%)	238 (47%)
Addresses reader	Negative	13 (3%)	46 (9%)
Voice	Positive	106 (21%)	118 (23%)
Aside to reader	Positive	20 (4%)	38 (8%)
Illegible handwriting	Negative	2 (0%)	7 (1%)
Poor spelling	Negative	52 (10%)	34 (7%)
Poor punctuation	Negative	102 (20%)	51 (10%)
Unfocused	Negative	31 (6%)	18 (4%)
Chain of ideas	Positive	4 (1%)	16 (3%)
Central idea	Positive	0	165 (33%)
Rhythm and flow	Positive	27 (5%)	35 (7%)
Inappropriate adjectives	Negative	0	13 (3%)
Onomatopoeia	Positive	0	13 (3%)

Correlation of Prominent-Feature Scores with NWP Analytic, Holistic Scores

Using all scores from the post writing assessment, correlations of individual prominent-feature scores with the NWP analytic and holistic scores were calculated and are presented in table 20. Because of the large number of cases ($N = 744$) on which these correlations were based, even very low correlations (e.g., absolute value of .10 or greater) would be statistically different from zero at the .01 level. Rather than interpret all such correlations, this section will summarize the patterns that may be discerned from table 20. (Very similar relationships were observed for the pre test scores, so they are not shown here.)

First, the correlations of individual prominent-feature scores tended to be very similar across the six analytic scores and the holistic score. There were some exceptions to this. Chief among these was the tendency for correlations of prominent-feature scores to be slightly lower with *conventions* scores than with any other of the NWP scoring categories. It is important to note, however, that such differences were not statistically tested.

Second, scores on prominent-feature elements considered to be positive attributes in an essay (e.g., balance/parallelism, voice) generally yielded positive correlations with the NWP scores, whereas scores on negative prominent-feature elements (e.g., weak structural core, poor spelling, unfocused) generally had negative or essentially zero correlations with the NWP scores. One exception was noted. The addresses reader scores correlated positively with several of the NWP scores, indicating that scorers may consider direct addresses to the reader a positive rather than a negative trait.

Third, the prominent-feature scores that showed the stronger relationships—in a relative sense, as none of the correlations observed were moderate or large—with the NWP scores were: (a) elaborated details, (b) dialogue, (c) sentence variety, (d) ending, (e) well-organized, (f) supporting details, and (g) voice. As was reported earlier, the overall prominent-feature scores correlated in the .40s with the NWP holistic score. Clearly, these prominent features (mostly positive) do appear to contribute to the scoring judgments on both the analytic and holistic sides.

Table 20. Correlations of Prominent-Feature Elements with NWP Scores for Combined Sites: Post Test (N = 744)

PF Element	Sentence						
	Content	Structure	Stance	Fluency	Diction	Conventions	Holistic
Voice	.32*	.30*	.34*	.27*	.30*	.22*	.33*
Elaborated details	.29*	.26*	.28*	.24*	.27*	.24*	.29*
Supporting details	.29*	.26*	.28*	.24*	.27*	.22*	.28*
Well-organized	.26*	.24*	.24*	.24*	.26*	.25*	.26*
Dialogue	.24*	.23*	.26*	.22*	.24*	.19*	.26*
Sentence variety	.22*	.20*	.23*	.23*	.22*	.24*	.22*
Central idea	.21*	.20*	.20*	.16*	.18*	.12*	.19*
End	.20*	.22*	.21*	.21*	.21*	.22*	.21*
Adverbial leads	.19*	.18*	.16*	.20*	.20*	.18*	.20*
Striking words	.18*	.15*	.17*	.18*	.20*	.15*	.19*
Balance/parallelism	.18*	.17*	.19*	.17*	.19*	.18*	.18*
Beginning	.17*	.20*	.18*	.19*	.19*	.18*	.18*
Chronological organization	.17*	.17*	.14*	.14*	.13*	.09*	.18*
Rhythm and flow	.17*	.17*	.17*	.16*	.19*	.16*	.19*
Aside to reader	.17*	.11*	.21*	.16*	.17*	.12*	.17*
Striking sentence	.16*	.14*	.18*	.16*	.20*	.13*	.17*
<i>Under-developed</i>	-.15*	-.14*	-.16*	-.17*	-.15*	-.16*	-.15*
Sensory language	.14*	.13*	.17*	.12*	.18*	.09	.15*
<i>Addresses reader</i>	.11*	.09	.12*	.06	.10*	.07	.10*
Cumulative sentences	.09	.08	.11*	.08	.17*	.07	.11*
<i>Poor spelling</i>	-.08	-.10*	-.08	-.12*	-.09	-.18*	-.09
Verb clusters	.03	.02	.05	.02	.10*	.02	.04
<i>Poor punctuation</i>	-.01	-.04	.00	-.05	-.00	-.10*	-.04
<i>Weak structural core</i>	-.02	-.06	-.05	-.10*	-.04	-.08	-.05
Supporting ideas	.09	.07	.05	.07	.08	.08	.07
Metaphors	.06	.04	.08	.05	.08	.05	.07
Repetition	.06	.06	.09	.05	.09	.04	.08
<i>Unfocused</i>	-.08	-.07	-.04	-.07	-.05	-.08	-.08
<i>Choppy</i>	-.06	-.06	-.06	-.06	-.05	-.04	-.08
<i>Usage problems</i>	.05	.05	.08	.04	.05	.02	.07
<i>Redundant</i>	-.02	-.02	-.05	-.03	-.00	-.00	-.02
<i>List of details/ideas</i>	-.01	-.01	-.03	-.03	-.02	.02	.00
<i>Weak organization</i>	-.01	-.01	.02	-.02	.02	.02	-.01

Notes: *Correlations having absolute value of .10 or greater are statistically significantly different from zero at the .01 level. Traits with less than 5% frequency are not shown. Elements considered negative appear in italics

Teacher Practices and Student Outcomes

How teaching practice is related to student outcomes is shown in table 21, a comprehensive table that shows the relationships among many of the complex factors in the present study. The teachers' scores are the Sum of Strategies score and the degree of implementation score. Student outcomes here are expressed in holistic writing score, the pre writing assessment given in the fall of 2005 and the post writing assessment given in the spring of 2006. The growth or the lack of growth is then expressed in the differences between these two scores—a reflection of what the teachers have taught over the course of the school year.

It is clear that the program school teachers tend toward a process model of teaching, while the comparison teachers tend toward a more traditional model. The mean sum-of-strategies score for the program group was 12.7 (on a scale of 0–20), compared to a score of 5.9 for the comparison group. On the degree of implementation score, the program group mean was 2.7 while the comparison group mean was 1.6 on a scale of 1–4.

Students in the program school showed a remarkable gain over the course of the school year, from 2.6 to 3.4 on a scale of 1–6, a mean gain of 0.8. On the other hand, scores of students in the comparison group remained flat, actually slightly regressing, - 0.1. It would appear that students in the comparison group learned very little about writing over the course of the year.

Table 21. Mean Implementation Score for the 10 Categories in the Degree of Implementation Continuum

Teacher	Grade Level	Years Experience	Strategies										Overall Strategy Score	Pre Test Score Holistic	Post Test Score Holistic	Amount Gain Holistic	Teacher Implementation Score	
			Student Choice Reading/Writing Connections	Prewriting	Peer Response	Teacher Conference	Mini-lessons	Revision	Editing	Publishing	Modeling							
Program School																		
P1	3	23	1	1	2	2	2	0	1	2	2	2	15	1.7	3.1	1.4	3	
P2	3	16	2	0	2	1	2	1	2	2	1	1	14	2.5	3.7	1.2	3	
P3	3	8	2	1	1	2	2	1	1	2	2	0	14	2.8	2.9	0.1	3	
P4	3	12	0	1	2	1	1	0	1	0	1	2	9	2.3	3.5	1.2	2	
P5	3	5	2	0	2	1	1	1	1	1	1	1	11	2.1	3.0	0.9	2	
P6	4	13	2	2	2	2	2	1	2	2	1	2	18	2.5	3.5	1.0	4	
P7	4	6	0	0	2	2	2	1	2	2	1	0	12	3.1	3.8	0.7	3	
P8	4	6	2	2	2	0	2	1	1	2	1	2	15	3.5	3.4	0.1	3	
P9	4	12	2	2	2	2	2	2	2	2	1	2	19	2.5	4.2	1.7	4	
P10	5	8	0	0	1	1	0	0	1	1	1	1	6	2.8	3.3	0.5	1	
P11	5	17	0	0	2	0	1	0	0	2	1	1	7	2.8	3.4	0.6	2	
Mean		11.5	1.2	0.8	1.8	1.3	1.5	0.7	1.3	1.6	1.2	1.3	12.7	2.6	3.4	0.8	2.7	
Comparison School																		
C1	3	8	0	0	0	0	1	0	0	1	1	0	3	1.8	2.2	0.4	1	
C2	3	33	0	0	1	1	1	0	0	1	1	0	5	2.3	2.5	0.2	1	
C3	3	7	2	1	2	1	0	1	1	2	1	0	11	2.4	2.8	0.4	2	
C4	4	35	1	0	0	0	0	0	0	0	0	0	1	2.8	3.1	0.3	1	
C5	5	4	1	0	2	1	2	0	0	2	2	0	10	3.0	3.3	0.3	3	
C6	4	24	1	0	1	1	1	0	0	1	1	0	6	3.8	2.8	1.0	1	
C7	5	7	0	1	0	0	0	0	0	1	0	0	2	3.6	2.4	1.2	1	
C8	5	10	0	1	2	2	0	0	1	1	2	0	9	3.1	2.9	0.2	3	
Mean		16.0	0.6	0.4	1.0	0.8	0.6	0.1	0.3	1.1	1.0	0.0	5.9	2.9	2.8	0.1	1.6	

Table 22 shows the relationship between the teachers' degree-of-implementation scores and mean sum-of-strategies scores, and the mean student gain in writing over the course of the year. It is clear that teachers with the highest degree-of-implementation scores also scored highest on the sum-of-strategies scores. It is also clear that students in the classrooms of teachers with the highest degree-of-implementation and sum-of-strategies scores outperformed students in more skills-based classrooms. Table 22 shows that the students of two teachers who scored at the highest implementation level, Level 4, improved 1.36 points on a six-point scale. At the other extreme, students of two teachers scoring at Level 1 and with the lowest sum-of-strategies scores remained flat or regressed. Between the two extremes, eleven teachers with degrees of implementation at levels 2 and 3 and with sum-of-strategies scores hovering around the median show student gains of about one-half to three-quarters of a point, not remarkable, but acceptable.

Although these results appear to be clear and dramatic, it should be kept in mind that this sample of teachers is rather small, only 19 in all, in both the program and comparison schools. One or two teachers could skew results in either direction in such a sample. Nevertheless the power of process teaching appears to be effective in promoting the growth of students' writing ability.

Table 22. Sum of Strategies Scores and Amount of Gain in Student Writing Based on Degree of Implementation Score for Program and Comparison Teachers ($N = 19$)

Teacher Degree-of-Implementation score	Frequency	Mean Sum-of-Strategies score (1–20)	Mean student gain (1–6)
4	2	18.5	1.35
3	7	12.7	0.49
2	4	9.5	0.77
1	6	3.8	-0.18

Overall, findings indicate that the professional development not only made a positive difference in the teaching practices of the program teachers, but also positively influenced the student writing in terms of number of words, T-unit length, presence of positive prominent features, and scores on both holistic and analytic measures.

Conclusions and Discussion

This study focused on writing achievement in grades 3, 4, and 5. Samples of student writing were taken in the fall of 2005, prior to the intervention, and again in the spring of 2006. Research focused on the features of that writing to analyze the amount of growth that occurred between writing samples. In this section, we discuss some of the conclusions and implications drawn from this study.

Effect on Teaching Practices

Although the program teachers were in their first year of writing project professional development, their teaching practices proved to be more process based and student centered than those of their comparison counterparts, despite similar educational backgrounds and teaching experience. For each of the strategies deemed important for the effective teaching of writing, the program teachers showed more complete implementation than the comparison teachers. Likewise, on the holistic degree of implementation score, program teachers scored higher than the comparison teachers. We believe, then, that the professional development had a profound effect on the practices of the program teachers. Further, the link between the professional development, teachers' practices, and the student outcomes in this study indicate that MWTI professional development for teachers had a positive effect on student outcomes.

Effect on Student Outcomes

Holistic Writing Scores and Student Outcomes

In overall holistic writing ability, the gains of the program group (0.7) are significantly greater than those of the comparison group (0.0). This indicates, first of all, that process teaching is superior to a traditional model for encouraging growth in writing ability, and it brings into question the efficacy of a traditional model: after a year in school, students in the comparison group overall showed no gain in writing ability. For those seriously interested in literacy learning, these findings both sound a warning and offer a glimmer of hope.

Analytic Writing Scores and Student Outcomes

Analytic writing scores include the six traits associated with writing assessment: *content*, *structure*, *stance*, *sentence fluency*, *diction*, and *conventions*. At the outset, the comparison group was superior in four of the six categories, and equal to the program group in the other two. On the post writing assessment, however, the program group proved to be significantly superior in all six categories; the difference was statistically dependable, and large enough to warrant attention.

In terms of *content*, the quality and clarity of ideas, the program group showed a significant gain compared to the comparison group. This gain may be seen as a measure of cognitive growth, as content development is related to understanding and expressing mental constructs, a crucial and important aspect of writing ability.

In the category of *structure*, defined as how effectively the writing establishes logical arrangement, coherence, and unity, the program group again significantly outperformed the comparison group. This category is deemed important because it relates to the skills of prioritization and categorization, and is applicable at every level of education.

The *stance* category describes how effectively the writing communicates a perspective through appropriate formality, style, and tone. Again writing growth of the program group proved significantly superior to that of the comparison group. In traditional instructional models, stance is not encouraged so much as adherence to rhetorical, grammatical, and usage restrictions. In a process model, stance is encouraged as the

starting place for writing, with focus on purpose and audience—including grammatical and rhetorical elements—following.

Sentence fluency refers to the production of a variety of sentence forms, and especially the modulation of sentence length. Over the course of the school year, the program group outperformed the comparison group. These young writers' emerging mastery of the sentence is encouraging. Sentence fluency can also be seen in light of the syntactic measure of mean T-unit length. Except for fourth-graders in the comparison group, the program group improved in T-unit length to a greater degree than their counterparts. (The fact that the small sample of fourth-graders in the comparison group increased their T-unit length more than the program group and yet did not improve their sentence fluency scores indicates the need for further study of the fourth grade papers of the comparison students.)

The primary factor in the *diction* category is the use of words and expressions appropriate for the context. *Diction* also refers to the use of word forms across parts of speech, including participle, adverbial, and adjectival forms to adapt words to fit the ideas. Metaphorical language, dialogue, and effective use of local expression are also related to quality of diction. The program group showed its greatest gain in this area. In the professional development, focus was on helping students produce diction appropriate to audience and purpose, rather than mechanically pulling multisyllabic synonyms from a thesaurus.

In punctuation, spelling, and capitalization, the program group again proved to be significantly superior. Conventions were not a singular focus of the professional development, but rather were taught at the point of need in classroom demonstrations. Thus, these gains can be considered tacit or residual, i.e., occurring in the context of other learning, but not directly taught as part of writing project professional development.

Syntactic Measures and Student Outcomes

Of the six standard syntactic measures (number of words, number of T-units, number of clauses, mean T-unit length, mean clause length, and subordination ratio), only two will be considered here: number of words, selected because previous research (Swain, Graves, and Morse 2006) has shown that length (number of words) correlates with higher assessment scores and mean T-unit length, selected because Kellogg Hunt's research showed that mean T-unit length is the most reliable indicator of syntactic maturity (Hunt 1965).

Number of words

The length of a piece of school writing has been shown to be positively associated with its quality (Swain, Graves, and Morse 2006), so it is important to look at this basic but important measure. Overall, for program students in grades 3, 4, and 5, the mean number of words grew from the pre writing assessment to the post writing assessment, though the fifth grade mean gain was slight. In the comparison school, the mean number of words declined from pre to post assessment. In view of the fact that only matched sets of papers were analyzed (i.e., the same students taking both the pre and post writing assessments),

we do not know how to account for this decline. The assessments were administered in the same fashion in both schools on both pre and post occasions, untimed and without interruption by loudspeaker or school activities. We assume that the comparison students became less engaged as writers during the year, while the program students became more engaged. This assumption indicates that a process approach to writing instruction may tend to keep students engaged as writers.

Mean T-unit length

It is important to note that no direct efforts were made to increase the young writers' mean T-unit length. That is, there were no sentence-combining exercises included in the professional development. However, several professional development sessions indirectly encouraged longer sentences, especially those sessions focused on the cumulative sentence form and on adverbial sentence leads, both of which serve to generate additional thought and elaborated detail. We assume that mean T-unit length will grow with professional development intervention, as well as with natural maturation. Therefore, where it does show growth—as in the program student data—we are encouraged. In any situation where mean T-unit length remains flat, and especially where it declines, the curriculum should be revisited.

Prominent Features and Student Outcomes

The correlations between prominent-feature scores and the holistic writing assessment scores as well as the correlations between prominent features and specific analytic writing traits underscore the relationship of prominent features to quality of writing. Furthermore, because the prominent features represent teachable writing strategies (as opposed to more abstract qualities such as a “sense of audience”), teachers can learn to guide their students to make use of positive prominent features in their writing. Further, writing project professional development can demonstrate for teachers how to teach the prominent features as part of the process of teaching students to write. Features such as dialogue, adverbial leads, cumulative sentences, asides to the reader, and striking words, to name a few, can and should become part of students' repertoire of writing tools for making their writing more effective.

Implications for Professional Development

This study indicates that through intensive and appropriate professional development, teachers can learn to teach young writers to enrich their writing through strategies that lead to more complete content, meaningful structure and organization, sophisticated sentence forms, and improved word choice and diction. Teachers can also learn to teach stylistic features that lend voice or stance to writing—and all this while continuing to maintain growth in use of conventions. Further, a pre analysis of the prominent features of students' writing, coupled with writing project professional development in the areas of need identified by the analysis, can result in multifaceted growth in the writing of elementary students, both in holistic writing ability and in the multiple traits of effective writing.

The present study points to the effectiveness of a professional development model that includes

- content influenced by student needs as determined by a prominent-feature analysis
- intensity of contact hours
- seminars
- classroom demonstrations
- off-site response to student writing
- multiple student revisions of developing papers.

Professional developers from the National Writing Project and across the nation can benefit from understanding the component parts of such a professional development model and considering how it might be adapted to benefit other schools and other young writers.

Bibliography/References

- Becker, A.L. 1965. "A Tagnemic Approach to Paragraph Analysis." *College Composition and Communication* 16: 237–242.
- Brooks, L. 1975. "The Effects of a Study of Generative Rhetoric on the Syntactic Fluency of Seventh Graders." *Unpublished Doctoral Dissertation*. Auburn, AL: Auburn University.
- Christensen, Francis. 1963. "A Generative Rhetoric of the Sentence." *College Composition and Communication* 14: 155–161.
- _____. 1965. "A Generative Rhetoric of the Paragraph." *College Composition and Communication* 16:144–156.
- Connors, Robert J. 2000. "The Erasure of the Sentence." *College Composition and Communication* 52: 96–125.
- Corbett, Edward P. J. 1991. *Classical Rhetoric for the Modern Student*. 3rd ed. New York: Oxford University Press.
- Crowhurst, Marion. 1980. "Syntactic Complexity and Teachers' Quality Ratings of Narrations and Arguments." *Research in the Teaching of English* 14 (3): 223–231.
- Crowhurst, Marion, and Gene L. Piche. 1979. "Audience and Mode of Discourse Effects on Syntactic Complexity in Writing on Two Grade Levels." *Research in the Teaching of English* 13: 101–109.

- Culham, Ruth. 2003. *6 + 1 Traits of Writing: The Complete Guide*. New York: Scholastic Professional Books.
- Elbow, Peter. 1994. "Writing First!" *Educational Leadership* 62 (2): 8–13.
- Faigley, Lester. 1979. "The Influence of Generative Rhetoric on the Syntactic Maturity and Writing Effectiveness of College Freshmen." *Research in the Teaching of English* 13: 197–206.
- _____. 1980. "Names in Search of a Concept: Maturity, Fluency, Complexity, and Growth in Written Syntax." *College Composition and Communication* 31: 291–300.
- Freedman, Sarah Warshauer. 1979. "How Characteristics of Student Essays Influence Teachers' Evaluations." *Journal of Educational Psychology* 71: 328–338.
- _____. 1987. "Summary of the Research." In *Response to Student Writing*, edited by Sarah Warshauer Freedman, Cynthia Greenleaf, and Melanie Sperling, 155–159. Urbana, IL: National Council of Teachers of English.
- Gebhard, Ann O. 1978. "Writing Quality and Syntax: A Transformational Analysis of Three Prose Samples." *Research in the Teaching of English* 12 (3): 211–232.
- Hillocks, George Jr. 1986. *Research on Written Composition: New Directions for Teaching*. Urbana, IL: ERIC Clearinghouse.
- _____. 2002. *The Testing Trap: How State Writing Assessments Control Learning*. New York: Teachers College Press.
- Hunt, Kellogg W. 1965. *Grammatical Structures Written at Three Grade Levels*. NCTE Research Report no. 3. Champaign, IL: National Council of Teachers of English.
- Huot, Brian. 1990. "The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends." *Review of Educational Research* 60 (2) 237–263.
- _____. 1996. "Toward a New Theory of Writing Assessment." *College Composition and Communication* 47: 549–566.
- Krishna, Valerie. 1975. "The Syntax of Error." *The Journal of Basic Writing* 1: 43–49.
- National Writing Project and Carl Nagin. 2003. *Because Writing Matters*. San Francisco: Jossey-Bass.
- Morse, David T., Sherry Swain, and Richard L. Graves. 2007. "Scaling Writing Proficiency via Prominent Features of Essays." A paper presented at the Annual Meeting of the Mid-South Educational Research Association, Hot Springs, AR.

- Nold, Ellen W., and Sarah W. Freedman. 1977. "An Analysis of Readers' Responses to Essays." *Research in the Teaching of English* 11 (2): 164–174.
- Northwest Regional Educational Laboratory (NREL). *Interactive Six Trait Writing Process*. Accessed July 7, 2006. Available at <http://senior.billings.k12.mt.us/6traits/>.
- O'Donnell, Roy C., William J. Griffin, and Raymond C. Norris. 1967. *Syntax of Kindergarten and Elementary School Children: A Transformational Analysis*. Champaign, IL: National Council of Teachers of English.
- Palacas, Arthur L. 1994. "Parentheticals and Personal Voice." In *Landmark Essays on Voice and Writing*, edited by Peter Elbow, 121–138. Davis, CA: Hermagoras Press.
- Roid, Gale H. 1994. "Patterns of Writing Skills Derived from Cluster Analysis of Direct-Writing Assessments." *Applied Measurement in Education* 7 (2): 159–170.
- Rubin, Donald L., and Gene L. Piche. 1979. "Development in Syntactic and Strategic Aspects of Audience Adaptation Skills in Written Persuasive Communication." *Research in the Teaching of English* 18 (4): 293–316.
- Sperling, Melanie, and Sara Warshauer Freedman. 2001. "Review of Writing Research." In *Handbook of Research on Teaching*, 4th ed., edited by Virginia Richardson, 370–389. New York: American Educational Research Association.
- Swain, Sherry, Richard L. Graves, and David T. Morse. 2005a. "The Effectiveness of Mississippi Writing/Thinking Institute Programs on Improving Student Writing Achievement and Teacher Practices at the Fourth Grade Level." Mississippi State University.
- Swain, Sherry, Richard L. Graves, and David T. Morse. 2005b. "The Effectiveness of Mississippi Writing/Thinking Institute Programs on Improving Student Writing Achievement and Teacher Practices at the Seventh Grade Level." Mississippi State University.
- Swain, Sherry, Richard L. Graves, and David T. Morse. 2006. "A Syntactic and Prominent-Feature Analysis of Seventh-Grade Writing." Provisionally accepted for publication by *Journal of Writing Assessment*, 2008.
- Witte, Stephen P., and Lester Faigley. 1981. "Coherence, Cohesion, and Writing Quality." *College Composition and Communication* 32: 189–204.

Appendix A. Interview Guide

Mississippi Writing/Thinking Institute Research Interview Form

MWTI Teacher Interview Guide

Interview

Evaluation Instrument for Implementation of Writing Project Professional Development

Teacher's Name _____ Cert. Level _____
 Name of School _____ N'tl Board Cert? _____
 School District _____ Years of Exp. _____
 Grade _____ Years in this school _____
 First yr. in this school: _____ Years in this grade _____
 Date _____

Part I

Use the student samples you brought with you to describe how you taught this piece of writing from beginning to end.

What decisions do you make along the way? How do you make those decisions?

Are there other ways you are helping your students learn to move through writing as process?

Initially, (at the beginning of the school year) how much time do you spend taking a piece of writing through a process to final draft?

What are the different reasons your students write?

Who are the audiences for their writing?

What are the different kinds of writing that your students do?

What strategies do you use to help students improve their papers?

How do you help students master grammar, mechanics, and usage?

Overall, what portion of your weekly teaching time with a single group of students is given to writing process activities or writing to learn in the content areas?

____ 0–24 percent

____ 25–49 percent

____ 50–74 percent

____75–100 percent

2. How often do you make use of large group, small group and individual instruction? What kind of instruction do you provide in each setting?

	Frequency	Kind of Instruction
Large group		
Small group		
Individual		

6. How do you know your students are improving as writers?

How do you use student writing to drive instruction? (Do you analyze student work?)

Give examples of purposes for which you use writing assessment results?

In the two or three weeks prior to the writing assessment, are there specific ways you address writing on demand?

If YES: Can you describe them?

7. Do you have students use writing as a tool for learning in content areas?

If YES: In what subject areas?

Do you formally collaborate with other content teachers to plan student writing?

If YES: Describe the collaboration and how you find the time to work with other teachers.

Do you make connections between reading and writing?

If YES: What strategies do you use? Can you describe one or two examples of making reading/writing connections?

8. A. How do you sustain your own professional growth as a writing teacher?

B. Are you a writer?

How do you sustain your own personal growth as a writer?

9. How have your administrators supported you in implementing strategies from the MWTI staff development?

What is your planning time like? Is there a common planning time?

How would you describe your school climate?

How have your peers supported you in implementing strategies from the MWTI staff development?

10. What additional comments do you have related to your participation in the MWTI staff development? (For comparison teachers: What additional comments do you have about your development as a writing teacher?)

Appendix B: Prompts

Prompt A

Different seasons: spring, summer, fall, winter, have different effects on our lives. Think about one season that affects you and the things you do. You might think about the weather or things that you do during this season, or both. Write to explain to a teacher how this one season of the year affects you. Support your ideas with examples and details.

Prompt B

Different kinds of weather such as sunny, stormy, windy, rainy, or snowy can affect our lives. Think about one kind of weather that had an effect on you or the people you know. You might think about how the weather condition changed the things you were able to do or not do. Write to explain to a teacher how this one kind of weather affected you. Support your ideas with examples and details.

Appendix C: Definition of Strategies

Definition of Strategies

- Evidence of student choice: The teacher deliberately plans for students to have choices of writing topics, reading materials, prewriting strategies, etc.
- Evidence of reading/writing connections: The teacher routinely helps students to connect what they are reading to what they are writing in terms of author's purpose, stylistics, use of language, etc.
- Prewriting: The teacher teaches a variety of prewriting strategies.
- Peer response: The teacher routinely relies on peer response to help students improve their writing.
- Teacher conference with student: The teacher routinely conferences with students to meet individual needs of all students.
- Mini lessons: The teacher designs brief lessons on style, sentence structure, language usage, grammar, word choice, or revision strategies based on student needs.
- Revision: The teacher incorporates time and strategies aimed at substantive revision of the content of student writing.
- Editing: The teacher addresses the grammar and mechanics of writing.
- Publishing: The teacher deliberately provides opportunities for students to publish outside the classroom.
- Modeling: The teacher models his/her own writing process or shares other models of writing with students as part of the writing instruction.

Appendix D: Continuum of Implementation

Descriptions of Practice at Each Score Point

[Remove or re-do TOC, jb!]

4 (Highest Degree of Implementation)

The teacher leads students through a series of recursive phases in writing, usually beginning with prewriting and concluding with publication activities; s/he adapts process to individual student needs.

The teacher engages in reflective practice, which includes analysis of student work, to drive instruction, articulating the goal of enabling students to connect to prior knowledge. Reflective practice informs the teacher of individual student growth. S/he plans individual or collective mini lessons to meet student needs.

The teacher routinely writes with students. S/he uses own writing as a model with students, allowing students to respond and revise. S/he writes and reads outside of class.

The teacher encourages thoughtful reading, rereading, and revision by providing a rich variety of group response to student writing, i.e., pairs, small group, large group, drama, tutorial or other audiences, as well as silent and oral reading of their writing, which results in multiple drafts.

The teacher welcomes an array of student-chosen or student-created types of writing. S/he coaches students through the writing phases with one-on-one and group conferences. S/he shares his or her own writing as well as examples of writing by students. The teacher offers students authentic publishing opportunities that reach multiple audiences.

The teacher routinely uses writing as a way of helping students deepen their understanding of various school subjects: literature, social studies, math, science, physical education, art, and music. The teacher articulates a well-developed and thorough understanding of the dynamics of writing to learn.

The teacher is a habitual reader and a writer. The teacher welcomes quality staff development and is able to articulate its value to his/her classroom. The teacher routinely attempts strategies demonstrated in staff development and makes adaptations for his/her class. S/he collaborates with other teachers and shares ideas and strategies, with or without an administrative structure for doing so.

3

The teacher leads students through a series of phases in writing. S/he partially articulates the reasoning behind the phases. The teacher sometimes engages in reflective practice, which may or may not include analysis of student work. S/he uses mini lessons for whole-class instruction.

The teacher sometimes writes and shares his/her own writing with students.

The teacher plans and provides some opportunities for revision, self-revision or teacher-assisted revision, and/or limited group response to student writing.

The audience sometimes extends beyond the school community; some student work is published. The teacher offers a variety of writing types.

The teacher sometimes uses writing as a learning tool, but not in a systematic or planned fashion. The teacher is not yet able to articulate a thorough understanding of writing to learn.

The teacher expresses enjoyment of reading, but may say s/he seldom finds the time to read. The teacher writes sporadically. The teacher comments on the value of staff development, attempts strategies presented, and sometimes, but not routinely, adapts and expands on the strategy to fit the needs of particular students. The teacher attempts to collaborate with other teachers, but is more likely to do so when there is a structure for collaboration in place.

2

The teacher assigns a topic or a group of potential topics for writing, and offers some help before collecting papers, grading, and returning them. The teacher rarely engages in reflective practice; s/he grades rather than analyzes student work. S/he may present mini lessons, but they are not tied to specific needs.

The teacher occasionally writes with students.

The teacher provides limited and random opportunities for revision, and does not articulate a plan for teaching revision.

The audience goes beyond the teacher, often to classmates or the principal, but there is no opportunity to publish beyond the school. Students have limited exposure to multiple types of writing.

The teacher randomly uses writing as a learning tool but with little or no understanding of the dynamics of writing to learn.

The teacher may or may not enjoy reading. His/her own writing is limited. The teacher may articulate a dislike for attending staff development and may admit to attending only for the Continuing Education Units. The teacher may attempt strategies demonstrated in staff development but when doing so makes no adaptations to accommodate his/her particular students. The teacher seldom voices a desire to work with other teachers.

1 (Little or No Implementation)

The teacher assigns a topic or a group of topics from which students write; offers little to no help; gives a deadline; collects papers; identifies errors; grades and returns papers to students. The teacher does not engage in reflective practice; he/she grades student work. He/she uses lecture or worksheet format, sometimes using process-oriented labels for text-based instruction.

The teacher does not write.

The teacher provides no opportunities for revision before the final, finished student writing is submitted to the teacher. The teacher controls correction in student writing from a predetermined list (e.g., agreement, spelling, and mechanics).

In most cases, the audience is the teacher, and there is little or no opportunity to publish. There is little exposure to various types of writing.

The teacher does not use writing as a tool for learning.

The teacher is not a reader or writer and may articulate resentment of any time spent in these activities. The teacher may articulate resentment about attending staff development, and may miss several sessions. There is no attempt to implement strategies from staff development. The teacher does not attempt to collaborate with other teachers.