NCEE 2012-4010

U. S. DEPARTMENT OF EDUCATION

An Investigation of the Impact of the 6+1 Trait Writing Model on Grade 5 Student Writing Achievement

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

An Investigation of the Impact of the 6+1 Trait Writing Model on Grade 5 Student Writing Achievement
Final Report

December 2011

Authors:

Michael Coe

Cedar Lake Research Group

Makoto Hanita

Education Northwest

Vicki Nishioka

Education Northwest

Richard Smiley

Education Northwest

Project Officer:

Ok-Choon Park

Institute of Education Sciences

NCEE 2012–4010

U.S. Department of Education

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Rebecca A. Maynard
Commissioner

December 2011

This report was prepared for the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, under contract ED-06-CO-0016 with Regional Educational Laboratory Northwest administered by Education Northwest.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the report.

This report is available on the Institute of Education Sciences website at http://ncee.ed.gov and the Regional Educational Laboratory Program website at http://edlabs.ed.gov.

Alternate Formats Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

Disclosure of potential conflict of interest

The design of this study along with all data collection, data analysis, and report writing were conducted by the Research and Evaluation Program of Education Northwest and by Cedar Lake Research Group, a subcontractor. None of the members of the research team have financial interests that could be affected by the content of this report. Professional development for teachers in the study was provided by the 6+1 Trait Writing Unit at Education Northwest; this unit does have financial interest that could be affected by the study but was not involved in the research activities.

Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

Acknowledgements

The authors would like to thank everyone who helped make this study possible, including the teachers, students, and administrators in participating schools; the technical advisors, reviewers, and editors who assisted us throughout the study; staff members at the U.S. Department of Education who were associated with the Regional Educational Laboratory Northwest during this contract; the Oregon Department of Education; and staff members at Education Northwest who contributed to the project.

Mike Puma and Dave Connell of Chesapeake Research Associates were invaluable advisers during all phases of the project, and provided a data management system that kept the research team blind to the origin of student essays during scoring. Andrew Jaciw of Empirical Education provided very helpful consultation on statistical issues as well as independent verification of the consistency and integrity of the data during the analytic process. Cris Price of Abt Associates helped resolve issues with the multiple imputation of missing data during the analysis. Carol Pistorino of Decision Information Resources, Inc. and her colleagues on the Analytical and Technical Support team, administered by Mathematica Policy Research, Inc., provided sound advice throughout the project. Peter Schochet of Mathematica was particularly helpful in resolving statistical modeling issues. Steve Slater at the Oregon Department of Education helped determine input parameters for the study design, based on Oregon statewide assessment data.

The REL Northwest Technical Working Group provided insight and guidance along the way. This group included Ray Barnhardt, Hans Bos, Audrey Champagne, Dave Connell, Bill Demmert, Allen Glenn, Dan Goldhaber, Brian Gong, Joan Herman, Michael Kamil, LeAnne Robinson, Lynn Santelmann, and Sam Stringfield.

The research team enjoyed the thorough and able support of Linda Fitch, Jennifer Klump, and Lisa Todd for information services; Matt Lewis for database management; Kathy Petersen, Tess Bridgman and Ann Rader for administrative support; and Bracken Reed and Nicole Sage for technical editing and communications. Management oversight and guidance was provided by Terri Akey, Bob Rayborn, Kim Yap, Steve Nelson, and Steve Fleischman. Art Burke and Richard Greenough provided statistical and technical advice. The student essays were scored by Linda Beith, Louise Brame, Adriane Burkhart, Dorthea Dober, Kerby Gault, Joan Hartley, Glenda Montgomery, Roger Rossman, Connie Spiegel, Ed Stites, and Marilyn Young, who were guided and coordinated by Sandra Marek.

Professional development services for teachers who participated in the study were provided by Jan Littlebear and Peter Bellamy with support from Mark Workman and Kay Dameron; Jan and Peter also provided much helpful input for our descriptions of the history, development and use of the 6+1 Trait Writing model. Gary Nave also provided assistance related to the description and implementation of the professional development. Ruth Culham of the Culham Writing Company provided historical insight into the development and key features of the 6+1 Trait Writing model.

Contents

Boxes

Figure

Tables

Summary

Reading, writing, and arithmetic have long been considered the foundation, or "basics," of education in the United States. Writing skills are important for an increasing number of jobs (National Commission on Writing 2004; Executive Office of the President 2009). Poor writing skills are a barrier to hiring and promotion for many individuals, and remediation of problems with writing imposes significant operational and training costs on public and private organizations (Casner-Lotto, Rosenblum, and Wright 2009; National Commission on Writing 2004, 2005). Writing is also important for the development of reading skills (Graham and Hebert 2010) and can improve learning in other academic content areas (Bangert-Drowns, Hurley, and Wilkinson 2004).

In response to the perceived neglect of writing in U.S. education, the National Commission on Writing proposed a set of recommendations for making writing a central element in school reform efforts (National Commission on Writing 2006). These concerns were echoed in regional needs assessment studies conducted by Regional Educational Laboratory (REL) Northwest, in which educators in the region placed a high priority on writing and literacy education (Gilmore Research Group 2006, 2009).

A growing body of research is beginning to shed light on classroom strategies and practices that improve the quality of student writing. For example, a recent meta-analysis of research on writing instruction in grades 4–12 finds support for 11 "elements of effective adolescent writing instruction" (Graham and Perin 2007a, 2007b). These recommended practices, synthesized from the findings of experimental studies, include having students analyze models of good writing; explicitly teaching students strategies for planning, revising, and editing their work; involving students in collaborative use of these writing strategies; and assigning specific goals for each writing project. These elements are core components of the intervention examined in this study.

The 6+1 Trait® Writing model (Culham 2003) emphasizes writing instruction in which teachers and students analyze writing using a set of characteristics, or "traits," of written work: ideas, organization, voice, word choice, sentence fluency, conventions, and presentation. The Ideas trait includes the main content and message, including supporting details. Organization refers to the structure and logical flow of the writing. Voice includes the perspective and style of the individual writer and his or her orientation toward the audience. Word Choice addresses the variety, precision, and evocativeness of the language. Sentence Fluency includes the rhythm, flow, and sound patterns in the construction of sentences that may make them pleasant and interesting to read. The Conventions trait, sometimes called mechanics, includes spelling, punctuation, grammar, capitalization, and other rule-based language forms. The trait of Presentation (the "+1" of the 6+1 Trait Writing model), which is focused on page layout and formatting issues, is related to the visual aspects of publishing writing. This trait might not be applied unless the writing project is carried through to publication or presentation in a classroom or

public forum. Presentation is not typically measured in large-scale assessments of student achievement, which require students to use particular formatting.

This framework and the associated terminology for characterizing the qualities of writing may be used to study the writing of others, to plan or revise one's own writing, or to discuss the qualities of a piece of writing with others. The 6+1 Trait Writing model includes many of the features recommended in the Graham and Perin (2007a) meta-analysis. This approach has been widely disseminated: the publisher of the model reports having distributed professional development materials in all 50 states and conducted professional development institutes or workshops in 48 states and several countries.

The model has not been adequately studied using experimental methods. In order to provide evidence on the effectiveness of this approach, the study reported here was designed as a large-scale effectiveness trial (Flay 1986). The model was not applied under ideal laboratory conditions, with frequent supervision by program developers to ensure optimal implementation. Instead, professional development was provided to teachers who worked in 74 Oregon schools that were randomly assigned to the study conditions. The professional development approach allowed teachers to implement the model in their classrooms according to their own style and preferences.

The study addressed the following confirmatory research question:

- What is the impact of 6+1 Trait Writing on grade 5 student achievement in writing?

It also investigated two exploratory research questions:

- What is the impact of 6+1 Trait Writing on grade 5 student achievement in particular traits of writing?
- Does the impact of 6+1 Trait Writing on grade 5 student achievement vary according to student gender or ethnicity?

As described further in Chapter 1, grade 5 students were chosen as the target population because the development of academic writing skills is key in this grade level—a time when students focus on learning expository and persuasive writing, which is used in much of their subsequent academic careers (Common Core State Standards Initiative 2010). Subgroup analyses by gender and ethnicity were deemed to be of interest because of the variation in student assessment outcomes based on these factors (Cole 1997; Nowell and Hedges 1998; U.S. Department of Education 2003a, 2003b).

Study sample and methods

Data for the cluster-randomized experimental study were collected from participating grade 5 teachers and students in 74 Oregon schools. Two cohorts of schools participated in the study across two consecutive years. The intervention and data collection occurred in 54 schools during 2008/09 and in an additional 20 schools in 2009/10. Schools were

first screened to ensure that they were not already using a trait-based writing instruction model, that they had an adequate number of grade 5 students to provide a reliable estimate of student performance, and that they were willing to participate. All procedures were the same in both cohorts of schools, and all data were combined for analyses (except for a specific analysis of cohort differences). Except where otherwise noted, this report describes the combined procedures and results of both cohorts of schools.

After administrators and teachers had been informed about the study and agreed to participate, each school was randomly assigned to either the treatment or control condition. Random assignment was done within pairs of schools that had been matched within each participating district to ensure that the treatment and control groups had similar percentages of students eligible for free or reduced-price lunch. (In districts with an odd number of participating schools, unpaired schools remaining after all paired schools were assigned were randomly assigned to the treatment or control condition.) Of the 74 schools in the study, 39 were randomly assigned to the treatment condition and 35 were randomly assigned to the control condition. As schools were the unit of random assignment, all participating grade 5 teachers in each school were assigned to the same condition.

Teachers in treatment group schools were offered training in the 6+1 Trait Writing model the summer before the data collection year and during that year. They learned how to apply the model and used it with students for the first time during the year in which student outcome data were gathered. Teachers in control group schools were not asked to change the instructional methods they would have used had they not participated in the study. The control condition thus represented a "business as usual" counterfactual in schools not already using trait-based writing instruction, with which the first-year implementation among treatment group schools was compared.

Teachers in both groups were asked to complete a survey at three points during the data collection year in order to report the extent to which they were using classroom practices recommended as part of the 6+1 Trait model. These self-report surveys were the only method used to determine whether treatment group teachers implemented the model with students or whether the practices recommended by the model were used by treatment group teachers more than they were used by control group teachers. Teachers in the two groups reported similar levels of use of these practices at the beginning of the study. By the end of the study, treatment group teachers reported greater use of these practices than did the control group teachers, but the newly developed survey instrument does not provide easily interpretable information about the magnitude of this difference or the specific level of implementation of these practices in treatment group or control group classrooms.

Within each school, all data for the study were collected during a single school year. Before the beginning of the year, teachers were asked to complete a questionnaire about their use of specific classroom writing instruction practices during the previous school year. Teachers in both the treatment and control groups reported that the classroom practices emphasized by the intervention were already in use at the outset of the study;

this was part of the existing school environment into which the intervention was introduced. Teachers in the treatment group then attended a three-day summer institute that provided comprehensive training, planning time, and resource materials to help them learn and apply the 6+1 Trait Writing model. During the following school year, teachers in treatment schools attended three additional one-day workshops to further their understanding of the approach and to plan trait-based writing activities for their students. Teachers in both the treatment and control groups completed a survey on classroom practices at midyear and again at the end of the school year.

Students in both control and treatment classrooms wrote essays at the beginning of the school year, which were scored and used as baseline measures of student writing performance. Students completed essays again at the end of the school year. Scores on this essay test were used as the outcome measures for the study. Each essay was rated using a single "holistic" score for overall writing quality. It was also rated on each of the six core characteristics of writing quality included in the 6+1 Trait Writing model. The holistic score was used for the confirmatory analysis and the second exploratory analysis; the trait scores were used for the first exploratory analysis.

Because the research team was employed by Education Northwest, the organization that developed and markets the 6+1 Trait Writing model, care was taken to maintain the transparency of all research processes and to limit the possibility of introducing intentional or unintentional bias at key phases of the research. The research team at Education Northwest was kept blind to key aspects of the data during the scoring of student essays, and they operated and were supervised independently of the individuals in a different organizational unit who provided the professional development. The teams of essay raters did not know whether a particular essay was a pretest or a posttest or whether it came from a treatment or a control school. Details of the methods used to prevent bias are provided in the report.

Summary of findings

The sample included 102 teachers and 2,230 students in the treatment condition and 94 teachers and 1,931 students in the control condition. The confirmatory research question was addressed by comparing the mean difference between posttest student essay scores in the two conditions, using a benchmark statistical model that accounted for students' baseline (pretest) writing performance at the beginning of the school year, the poverty level of their school, and preexisting baseline differences between schools on three teacher-reported characteristics: the school average for the weekly teacher-reported hours students spend in class practicing writing, the school average for teacher years of teaching experience, and the school average for teacher years of experience teaching writing. The statistical model also took into account the fact that students were clustered within schools and therefore were more likely to be similar to one another than would have been the case had students rather than schools been randomly assigned to conditions. Following a plan defined prior to implementing the study, the benchmark estimates of effectiveness were based on a statistical analysis that imputed the outcome measures in cases where they were missing (5.5 percent of all cases). The effectiveness of

the professional development was also estimated using a more common approach of deleting cases with missing values of the outcome measure. Another alternative analysis used a different statistical model to examine the data.

The benchmark estimates indicate that use of the 6+1 Trait Writing model significantly increased student writing scores during the year in which it was introduced to schools. After controlling for baseline writing scores, the estimated average score of students in the treatment group was 0.109 standard deviations higher (p = .023) than the estimated average score of students in the control group. An intervention with this effect size would be expected to increase the average level of achievement from the 50th to the 54th percentile.

The findings remained stable when tested using alternative choices for the analytic sample and the model specification. When students with missing data were excluded from the analysis sample, the estimated effect was 0.110 standard deviations (p = .018). Use of an analytic model that did not adjust for baseline measures of teacher experience and instructional practices resulted in an estimated effect size of .081 (p = .048).

In addition to the analysis of holistic writing scores, exploratory analyses found statistically significant differences between control and treatment group students on three of the six specific outcome measures of particular writing traits — organization, voice, and word choice — with effect sizes ranging from 0.117 to 0.144 (p = .031 to .018). For the other three traits — ideas, sentence fluency, and conventions — the mean outcome score of students in the treatment condition was higher than that of students in the control condition, but these differences were too small to be considered statistically significant given the size and sensitivity of the experiment.

Additional exploratory analyses of holistic writing scores found no differential effects of the intervention based on student ethnicity or gender.

Limitations

The findings reported here are limited by several contextual factors:

- The intervention studied in this research was a first-year implementation of the 6+1 Trait Writing model, which provided additional writing instruction and assessment strategies that were intended to complement whatever writing curricula and instructional strategies were present in the participating schools. Questions about the interaction of the model with any specific writing curriculum were not addressed and cannot be answered using these findings. Questions about curriculum materials designed to fully integrate a trait-based approach to writing were not addressed by this research; the findings presented here cannot be applied to answer such questions.
- The implementation of recommended classroom strategies by the treatment group and control group teachers was measured using newly developed self-report surveys that have not been validated by observational or other measures. These surveys provided

information about implementation by one group relative to the other group and were subject to possible biases in teacher self-reports of their classroom practices. The extent to which the model was actually implemented by treatment group teachers is unknown, as is the extent to which treatment group teachers implemented these strategies more than they were implemented by the control group teachers.

- The findings reported here are for grade 5 students in 74 Oregon schools that volunteered to participate. The extent to which these findings may apply to other grade levels, other schools, or other regions is unknown.
- The extent to which the findings would be replicated in other settings, such as targeted implementations for particular schools or student populations, is unknown and cannot be inferred from these results.
- The student achievement data were collected during the same school year in which teachers received their first year of professional development in the 6+1 Trait model. The study does not answer questions about what effects might be produced by longer durations of professional development and/or classroom implementation.
- It is possible that teachers or students in the treatment group may have responded differently to the knowledge that they were participating in an experimental study than did teachers or students in the control group; if so, any difference or lack of difference in the performance of teachers or students in the two groups could have been due in part to this differential response to participation in a research study.

1. Study background

Writing ability is a critical component of language and literacy, with a complex relationship to reading ability (Fitzgerald and Shanahan 2000, Graham and Hebert 2010) and to learning and thinking within other disciplines (Bangert-Drowns, Hurley, and Wilkinson 2004; Keys 2000; Shanahan 2004; Sperling and Freedman 2001). Writing is an essential skill not only for academic development but also for success in an increasing number of occupations (National Commission on Writing 2004; Executive Office of the President 2009). Despite the historic emphasis on "reading, writing, and arithmetic" in U.S. schooling, the development of writing skills has received less emphasis in policy and practice than the development of student skills in reading or mathematics (National Commission on Writing 2003, 2004).

Many students in the United States do not receive adequate education in writing to prepare them for the workplace. In 1993 the U.S. Department of Labor identified writing as one of the foundation skills required for successful employment in a broad range of jobs (U.S. Department of Labor 1993). Since that time, the proportion of U.S. jobs that require postsecondary education and/or written communication skills has grown (National Commission on Writing 2004; Executive Office of the President 2009). However, in 2002 and again in 2007, the National Assessment of Educational Progress (NAEP) found that 24 percent of students in grade 12 were proficient in writing (Persky, Daane, and Jin 2003; Salahu-Din, Persky, and Miller 2008). Almost a third (32 percent) of high school graduates taking the ACT are not adequately prepared for college composition courses (ACT 2005). In fall 2000, 23 percent of freshmen at public two-year colleges and 9 percent of freshmen at public four-year colleges enrolled in remedial writing courses (Parsad and Lewis 2003). In 2003/04, 10 percent of all students at public two-year colleges and 8 percent of all students at public four-year colleges enrolled in remedial writing courses (Berkner and Choy 2008).

In 2006, 81 percent of a sample of 431 corporate human resource professionals and senior executives listed applied skills in written communication as being deficient among U.S. high school graduates (Casner-Lotto and Barrington 2006). Writing was the area of greatest deficiency noted for both applied and basic skills among recent high school graduates. Among basic skills, more survey respondents (72 percent) cited deficits in writing skills than deficits in mathematics (54 percent) or reading (38 percent). Corporations and state governments report that poor writing skills are a barrier to hiring and promotion for many individuals and that remediating problems with writing imposes significant operational and training costs on their organizations (Casner-Lotto, Rosenblum, and Wright 2009; National Commission on Writing 2004, 2005). In response to the perceived neglect of writing in U.S. education, the National Commission on Writing proposed a set of recommendations for making writing a central element in school reform efforts (National Commission on Writing 2006).

A growing body of research is beginning to shed light on classroom strategies and practices that improve the quality of student writing. For example, a recent meta-analysis of research on writing instruction in grades 4–12 finds support for 11 "elements of effective adolescent writing instruction" (Graham and Perin 2007a, 2007b). These recommended practices, synthesized from the findings of experimental studies, include having students analyze models of good writing; explicitly teaching students strategies for planning, revising, and editing their work; involving students in collaborative use of these writing strategies; and assigning specific goals for each writing project. These elements are core components of the intervention examined in this study.

Need for the study

In 2006 the Regional Educational Laboratory (REL) Northwest conducted a series of hearings to obtain input into the types of evidence educators need to guide policy and practice decisions. Literacy, including both reading and writing, was identified as a critical issue in the region. A theme of the discussions was the need for evidence about effective practices for writing instruction (Gilmore Research Group 2006). In 2007 and 2008, these hearings were followed with a needs assessment survey of educators in the Pacific Northwest region, with the goal of establishing "a road map by which to plan programs and set a meaningful research agenda to address state and regional educational needs" (Gilmore Research Group 2009, p. 1). Improving student literacy was identified as a high priority among educators in Alaska, Idaho, Montana, Oregon, and Washington. When asked to prioritize their needs for evidence related to improving literacy instruction by allocating points to research focused on potential topics, "integrating reading and writing across the curriculum" received the highest priority ranking among superintendents, principals, and teachers, with 31–36 percent of each of these groups assigning the integration of reading and writing instruction 4 or more points out of 10 (Gilmore Research Group 2009).

Student performance on the National Assessment of Educational Progress (NAEP) in the region illustrates the need for improvement in student writing. Grade 4 students in Idaho, Montana, and Oregon performed below national levels on the 2002 NAEP examinations: 22 percent of students in each state scored at or above proficiency, well below the national public school rate of 27 percent. The rate for Washington students was 30 percent; Alaska did not participate (Persky, Daane, and Jin 2003). Grade 4 students were not tested in writing during the 2007 NAEP assessment (Salahu-Din, Persky, and Miller 2008). Among public school students in grade 8, the national proficiency rates in writing were 30 percent in 2002 and 31 percent in 2007; rates in Pacific Northwest states were 29–34 percent in 2002 and 29–35 percent in 2007. (Alaska did not participate in either year of testing; Oregon did not participate in 2007.)

The goal of this study is to provide high-quality evidence on the effectiveness of the 6+1 Trait Writing model for increasing student achievement in writing. Trait–based writing instruction is based on the use of a set of rubrics (scoring guides) to describe and assess different characteristics of an essay or other written work. For example, the ideas presented in an essay may be considered separately from the way the essay is structured

or organized; the use of standard conventions or mechanics, such as punctuation and spelling, may be considered another distinct feature or trait of the essay. A focus on specific traits of writing may give students and teachers a shared framework and vocabulary to identify and discuss strengths and weaknesses while allowing them to form a plan to revise a particular essay or to build skills in a particular aspect of writing. One student's essay might include interesting and innovative ideas, presented with perfect spelling and punctuation but in a disorganized logical flow. Another student might write an essay that includes interesting ideas presented in a well-organized format but that is marred by spelling and punctuation errors.

When using a scheme for separating the traits of a written work in this way, teachers may use an assessment "rubric," or scoring guide, to give students detailed, systematic feedback on the different traits of their writing, as well as targeted suggestions for improvement. The delineation of a system of traits provides a common structure and vocabulary with which students and teachers can think about and discuss their writing. This trait-based method of planning, assessing, and revising writing (sometimes called an "analytic" method) is an alternative to "holistic" assessment and feedback, in which a single score is given to a student essay.

The publisher of the 6+1 Trait Writing model, Education Northwest (which also houses REL Northwest, the research group that conducted this study), reports distributing training materials for this model or previous versions of the model to districts in all 50 states and internationally and providing professional development sessions in 48 states and several countries. Similar trait-based approaches to writing instruction have been incorporated into language arts curriculum materials and guides to writing instruction that are distributed through major educational publishers, training workshops, websites, and other resources for schools. Several states use their own variant of a trait-based approach for their educational standards and/or statewide assessment systems for student writing skills. In an unpublished 2009 review of state education standards and assessments, the developer of the 6+1 Trait Writing model found that at least 22 states had some variant of trait language in their writing standards and at least 35 states included some variant of the writing traits in their writing assessments. (These counts do not include standards or assessments related to the trait of "conventions," which refers to standard usage of punctuation, spelling, grammar, and capitalization. These are included in some form in all state standards and assessments for writing.)

Given the importance of writing for academic and career success and the widespread distribution of trait-based writing instructional materials and assessments, it is important that the education community have access to high-quality scientific evidence on the effectiveness of the trait-based approach. This study contributes to that knowledge base by helping to ensure that decisions about the adoption of this approach are based on reliable evidence. As detailed in the next section, the particular writing model studied here, 6+1 Trait Writing, contains many of the elements of writing instruction that have been found to be effective in past research. However, this particular combination and implementation of these practices has not been previously studied using rigorous methods of research. Other trait-based writing instruction materials have also been available at

times from ad hoc sources such as educator websites or as part of published curriculum and instruction materials, but the widespread use of the 6+1 Trait Writing model and the consistent availability of 6+1 Trait materials and professional development for grades K–12 led to the choice of this particular model for the study.

An overview of 6+1 Trait Writing

"Six-trait writing" was developed in the 1980s as an approach to classroom assessment of student writing that would provide students and teachers with more structure to understand how to write well (Culham 2003). The method was built on the descriptive and theoretical work of Diederich (1974) and Purves (1988), two pioneers in the use of classroom-based analytical assessments of student writing to inform diagnostic decisions about writing instruction. It incorporates aspects of process writing, including the recursive use of planning, drafting, assessment, and revision developed through the work of Emig (1971), Flower and Hayes (1981), Applebee (1986), and others.

Early development of the 6+1 Trait Writing model was also informed by the work of George Hillocks. In a meta-analysis of studies on writing instruction, Hillocks (1986) examined six types of instructional focus, including an emphasis on grammar, a focus on studying models of good writing, the use of "sentence combining" to gain fluency with different kinds of sentence structures, the use of "scales" to "present students with sets of criteria for judging and revising compositions," practice with "inquiry" methods to learn strategies for using data in their writing, and "free writing," also referred to as "the process approach to writing." Hillocks concluded that a focus on grammar was actually detrimental to student progress in writing but that the other five instructional focuses yielded positive effect sizes. The weakest effects were associated with free writing and focusing on models; the strongest effects were associated with a focus on inquiry, scales, and sentence combining. Hillocks concluded that "free writing and the attendant process orientation are inadequate strategies" on their own and should be combined with an explicit focus on sentence structures, manipulation and organization of information into coherent arguments or narratives, and use of specific criteria to assess and revise writing in a recursive fashion. (See Hillocks 1987 for a succinct discussion of this study and its implications for writing curricula.)

The materials that became the foundation for the 6+1 Trait Writing model were developed by teachers in Oregon and Montana, based on work by Diederich (1974), who identified five characteristics of writing during his examination of detailed reviews of student writing. The six-trait assessment and feedback model included a set of writing characteristics that were somewhat different from the five characteristics of student writing that had been proposed by Diederich; their development was supported through funding by the U.S. Department of Education to the Northwest Regional Educational Laboratory (now reorganized as Education Northwest). These materials were not placed under copyright; educators' and publishers' freedom to copy or adapt them led to a proliferation of six-trait materials that have since been formally published or informally shared among educators.

In 1999 the scoring guide for the traits was revised and a seventh trait was added to address visual formatting of student writing for publication or presentation at public events. Tools for publication and presentation were becoming more available to students through desktop publishing and presentation software. Because the use of these tools may not be appropriate in all educational settings or applications, the new seventh ("+1") trait was considered to be an important but optional part of the model. Updated materials were placed under copyright using the name "6+1 Trait Writing."

The 6+1 Trait Writing model can be used to provide additional structure, content, and guidance to instruction that is based on the writing process (or "writing workshop") format, which began as a method that gave students little specific guidance on writing. According to Pritchard and Honeycutt (2006, p. 275), "The understanding of what constitutes the writing process instructional model has evolved since the 1970s, when it emerged as a pedagogical approach. In the early years, it was regarded as a nondirectional model of instruction with very little teacher intervention." Although current process writing instruction may include a prescriptive structure involving planning ("prewriting"), writing, and rewriting, these steps may be applied in a formulaic manner that does not focus attention on the interplay between intended audience, author's voice, ideas and organization, word choice, and sentence fluency (Boscolo 2008). Writing instruction may be enriched by an explicit focus on particular traits of writing.

Some cognitive developmental approaches to writing instruction attempt to break down the complex task of writing by focusing on discrete tasks through the use of worksheets and exercises, which may be used in general classroom writing instruction or in writing instruction for special education populations or students with particular difficulties in writing (Scardamalia, Bereiter, and Fillion 1981; Harris and Graham 1996). The concepts and language of trait-based writing can be integrated with these approaches to provide more context and definition to the discrete tasks and strategies that are taught during these classroom activities. This framework can also help teachers in different classrooms and grade levels communicate with students about writing using a common set of expectations and descriptors (Culham 2003).

The 6+1 Trait Writing model is not an alternative writing curriculum designed to replace existing writing programs in schools, but rather an additional, complementary set of tools to aid in conceptualizing, assessing, and describing the qualities of writing. It is used in conjunction with existing writing curricula to provide a framework for classroom writing instruction, feedback, and dialogue that is designed to improve the ability of K–12 teachers and students to plan, evaluate, discuss, and revise their writing (Culham 2003).

The model includes a framework of instructional *strategies* (classroom practices) that are used to facilitate the integration of assessment with instruction, targeting seven *traits* of effective writing: ideas, organization, voice, word choice, sentence fluency, conventions, and presentation (box 1; Culham 2003). The 6+1 Trait Writing model can be used within language arts instruction or for applications in which writing is integrated with other academic subjects. Because the model is intended to be integrated with existing writing

curricula and instruction, rather than replacing them, teachers have flexibility in how this is done, including the intensity with which the suggested practices are used with students.

Box 1. The 6+1 Traits

1. Ideas

Ideas are the main message, the content of the piece, the theme, together with the details that enrich and develop that theme.

2. Organization

Organization is the internal structure, the thread of central meaning, the logical and sometimes intriguing pattern of the ideas within a piece of writing.

3. Voice

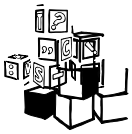
Voice is the heart and soul, the magic, the wit, along with the feeling and conviction of the individual writer coming out through the words.

4. Word Choice

Word choice is the use of rich, colorful, precise language that moves and enlightens the reader.

5. Sentence Fluency

Sentence fluency is the rhythm and flow of the language, the sound of word patterns, the way in which the writing plays to the ear and not just to the eye.



6. Conventions

Conventions refer to the mechanics of writing: spelling, paragraph formatting, grammar and usage, punctuation, and use of capitals.

+1 Presentation

Presentation zeros in on the form and layout of the text and its readability; the piece should be pleasing to the eye.

Source: 6+1 Trait Writing Summer Institute training agendas and records.

In the 6+1 Trait Writing model, teachers are provided with a range of activities to support classroom instruction on the writing traits and to engage students in learning about and practicing the use of the traits in planning, assessing, and revising their writing. The 10 instructional strategies that provide the framework for writing instruction in the 6+1 Trait Writing model are listed below:

1. Teaching the language of rubrics for writing assessment.
2. Reading and scoring papers, justifying the scores, and having students do so themselves.
3. Teaching focused revision strategies.
4. Modeling participation in the writing process.
5. Having students read and analyze materials that demonstrate varying writing quality.
6. Giving students writing assignments to respond to effective prompts (that is, prompts that are engaging to students and provide adequate structure, guidance, and context to elicit detailed responses).
7. Weaving writing lessons into other subjects.
8. Teaching students to set goals and monitor their progress.
9. Integrating learning goals for writing into curriculum planning.
10. Teaching ways to structure nonfiction writing.

More detail on these strategies is provided in box A1 in appendix A.

A key component of the 6+1 Trait Writing model is the ongoing use of rubrics (scoring guides) to provide feedback to students on their writing. Feedback is provided through self-scoring, peer scoring, and teacher scoring of student writing at various stages of the writing process (planning, drafting, revision, redrafting). In addition to the studies of writing instruction described above, experimental and correlational studies of subjects other than writing instruction find positive associations between student performance and the use of formative assessment and feedback as strategies to support student learning (Black and Wiliam 1998; Crooks 1988; Marzano 2003; Natriello 1987; Sadler 1989).

Recent development of the 6+1 Trait Writing model has focused on classroom instruction and the use of the traits to help students plan and revise their writing, through the use of the writing trait rubrics for formative assessment (self-assessment, peer assessment, and teacher feedback) on student writing. Writing assessments based on the 6+1 Trait Writing model could potentially be used to screen students at risk of failure on statewide assessments of writing performance. In a correlational study of the correspondence of six-trait writing scores and performance on the Washington Assessment of Student Learning (WASL) writing test, Coe (2000) finds that students who scored low on the six traits tended to also score low on the WASL. For example, 28.6 percent of students who had at least one trait score less than 3.0 (using a five-point scale) passed the WASL. Conversely, 83.1 percent of students with scores of 3.0 or above on all six traits passed the WASL. Among students with all trait scores above 3.5, 93.8 percent passed the WASL writing test.

A recent meta-analysis of writing instruction strategies finds experimental evidence for several of the core instructional activities incorporated in the 6+1 Trait Writing model (Graham and Perin 2007a, 2007b). Table 1 illustrates correspondences between the 11 strategies recommended by Graham and Perin (2007b) in their report *Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High Schools* and specific instructional strategies and student activities included in the 6+1 Trait Writing intervention.

Table 1. Elements of 6+1 Trait Writing that correspond to recommendations in *Writing Next*

Writing Next recommended elements of effective writing instruction	6+1 Trait Writing activities
Writing strategies	Emphasizes use of writing traits for planning, revising, and editing compositions
Summarization	Includes use of writing traits to analyze and summarize texts
Collaborative writing	Encourages collaboration in planning, drafting, revising, and editing
Specific product goals	Assigns specific goals for writing and then, using trait rubrics, helps students routinely self-assess
Word processing	Encourages the use of appropriate technology to support students in the development and publication of compositions
Sentence combining	Teaches students to understand and construct more complex, sophisticated sentences by combining, rearranging, expanding, and imitating sentences
Prewriting	Encourages students to generate, gather, and organize ideas for their compositions before writing
Inquiry activities	Encourages students to organize and analyze information to develop content for their writing
Process writing approach	Includes instructional activities appropriate for a process writing workshop environment; graphic overviews show how the model works with the writing process
Study of models	Encourages students to read and analyze models of good writing using the writing traits, focusing on particular audiences, purposes (modes), and forms
Writing content learning	Integrates writing with content in other subjects

Source: 6+1 Trait Writing model training materials.

The professional development offered to teachers as part of this study included a three-day summer institute and three one-day workshops during the school year intended to help teachers learn about 6+1 Trait Writing and implement the approach with their students. Teachers also had access to online support resources to help them integrate the model into their existing classroom writing instruction.

Previous research on 6+1 Trait Writing

Little research focuses specifically on the effectiveness of the 6+1 Trait Writing model for classroom writing instruction. Two experimental research studies examine the impact of earlier versions of the 6+1 Trait Writing model on student learning (Arter et al. 1994; Kozlow and Bellamy 2004). However, technical flaws in these studies limit their value for understanding the impact of the intervention, highlighting the need for more rigorous research. Moreover, the 6+1 Trait Writing model has been developed further since these studies were conducted, limiting their usefulness for understanding the effectiveness of the current intervention.

The first study was conducted by Arter and colleagues during the 1992/93 school year in three schools, including six grade 5 classrooms that were randomly assigned to either a treatment or control condition. Teachers in the treatment group received trait-based training and specific lesson plans and strategies to enhance student writing performance, followed by technical assistance during the year. Teachers in control classrooms continued writing instruction as usual. Students completed pretest essays at the beginning of the year and posttest essays at the end of the year, which were scored by raters on six writing traits. On the Ideas trait, the scores of students in the treatment group increased by more than the scores of students in the control group, a difference that was reported to be statistically significant. Gains in other specific traits were also larger among treatment group students but were not reported to be statistically significant. Effect sizes and related variance estimates were not reported.

The results of this study must be interpreted with care because the analysis did not properly account for the nested data structure (the fact that students within the same classroom are likely to have similar scores). This violates the assumptions of the statistical analysis that was used and may have resulted in differences between groups being erroneously identified as statistically significant. The raw data no longer exist, making reanalysis using proper statistical adjustments impossible. Details of the research design and implementation are no longer available, making it impossible to judge the extent to which appropriate standards of experimental design were applied during this study.

The second study, conducted by Kozlow and Bellamy during the 2003/04 school year, involved 72 classes in grades 3–6 in one school district. Within each grade level, half of the district's teachers were randomly assigned to the treatment group and half to the control group. (Randomization was not completely preserved, however, because teachers who were originally assigned to the treatment group but could not attend the training were reassigned to the control group.) Teachers in the treatment group received two days

of training in November and were asked to incorporate the trait-based approach as part of their classroom teaching and assessment of student writing for the remainder of the school year. Teachers in the control group were asked to continue teaching and assessing student writing using their normal practices. Classroom visits showed considerable variation in the extent of implementation by teachers in treatment groups, as well as substantial implementation of similar practices in the control group. Students completed pretest essays at the beginning of the year and posttest essays at the end of the year, which were scored by raters using both a single holistic rating of writing quality and six separate ratings for specific writing traits. The data were analyzed using a mixed-model framework to account for the nested data structure. Differences between the treatment and control group were not statistically significant. The amount of professional development offered to teachers was less than that currently recommended by the developers and less than the amount tested in this report. Teacher survey comments indicated a need for more training and more time to implement the model.

These studies do not provide reliable evidence regarding the effectiveness of the 6+1 Trait Writing approach to writing instruction. Both experimental studies had serious flaws by current standards, including improper statistical analyses, failure to preserve the integrity of their randomized designs, and sample sizes that yielded inadequate statistical sensitivity for the detection of a treatment effect.

Need for experimental evidence on 6+1 Trait Writing

The randomized controlled trial reported here is an important contribution to regional and national educational research for the following reasons:

- Writing is a key skill for academic and career success but has received less emphasis in both research and policy than reading and mathematics.
- The 6+1 Trait Writing model and related trait-based approaches to writing instruction and assessment are widely used and embody practices that have been supported by both the theoretical literature and empirical studies.
- No high-quality experimental studies have been conducted to estimate the effectiveness of the 6+1 Trait Writing model.
- Well-designed randomized controlled trials provide the best estimate of the impact of education interventions on student achievement.

Research questions

This study was designed to answer one confirmatory and two exploratory research questions. The experiment was intended primarily to determine the impact of the intervention on student writing achievement during the first year of implementation, under conditions that would be typical for teachers receiving 6+1 Trait Writing professional development. The confirmatory research question was:

- What is the impact of 6+1 Trait Writing on grade 5 student achievement in writing?

Grade 5 was chosen as the target population in order to test the intervention at a grade level in which the demand for expository, academic writing is increasing, while avoiding pragmatic conflict with the Oregon Assessment of Knowledge and Skills statewide student assessment in writing, which is administered in grades 4 and 7. In 2005, at the time of the study design, Oregon standards for student expository writing rose from a recommended production of 100 words in grade 3 to 400 word essays in grade 5, with increasing levels of sophistication similar to those of the recently developed Common Core State Standards for language arts, which have since been adopted by Oregon (Common Core State Standards Initiative 2010). As noted in the background material for the Common Core State Standards, writing instruction in K–12 schools, as embodied in the NAEP assessment framework, includes a mix of three "mutually reinforcing writing capacities: writing to persuade, to explain, and to convey real or imagined experience." As students progress through the grade levels, writing to convey personal experience becomes progressively less heavily weighted and students are expected to produce more persuasive and expository writing (Common Core State Standards Initiative 2010, p. 5). These standards also emphasize the importance of the writing-reading connection and the "centrality of writing to most forms of inquiry," making the integration of research and writing skills a key focus of the standards (Common Core State Standards Initiative 2010, p. 8).

Two exploratory research questions addressed whether there were different impacts on measures of particular writing traits and whether the impacts depended on gender or ethnicity. These exploratory questions were:

- What is the impact of 6+1 Trait Writing on grade 5 student achievement in particular traits of writing?
- Does the impact of 6+1 Trait Writing on grade 5 student achievement vary according to student gender or ethnicity?

Subgroup analyses by gender and ethnicity were deemed to be of interest because of the variation in student assessment outcomes based on these factors. For example, on the Oregon statewide writing assessment, girls in grade 4 had proficiency rates in writing that were 15.7 percent higher than boys in 2006/07 and 14.8 percent higher than boys in 2007/08. Similarly, NAEP writing assessment data from Oregon in 2002 show a 19-point scale-score gap between boys and girls in grade 4 writing and a 23-point gap between boys and girls in grade 8 writing (U.S. Department of Education 2003a, 2003b). For comparison, the grade 4 gap between free or reduced-price lunch eligible students and those not eligible was also 19 scale-score points and at grade 8 it was 27 points. The gender gap is similar in size to the gap based on students' family incomes. These findings mirror persistent differences between boys and girls observed nationally, in which the gap between girls and boys in writing proficiency was greater than in any other academic subject and remained stable between 1960 and 1990 (Cole 1997; Nowell and Hedges 1998).

Because the research team was employed by Education Northwest, the organization that developed and markets the 6+1 Trait Writing intervention, specific methods—described in chapter 2 of this report—were used to maintain the transparency of all research processes and limit the possibility that intentional or unintentional bias could be introduced at key phases of the research. These arrangements included the following key elements:

- The Education Northwest research team that conducted the evaluation was made up of individuals who were employed within a separate organizational unit, distinct from the unit that publishes and disseminates the intervention.
- Randomization of schools to experimental condition was performed by an independent contractor that had no specific knowledge of the schools or districts involved, to prevent any possibility that such knowledge within Education Northwest could influence the randomization process.
- Student essays were transmitted directly from schools to the same independent research firm, which masked the origin of the essays by assigning new identification numbers known only to them before sending copies of the essays to Education Northwest for scoring. During scoring, the teams of essay raters at Education Northwest did not know whether a particular essay was a pretest or a posttest or whether it came from a treatment or a control school.
- After the essays were rated, the scores were transmitted back to the independent subcontractor, who merged these scores with the student demographic information and restored the original identification numbers linking the essays to particular schools and test administrations. This data file was used for the analysis.
- All subsequent manipulations of the data file (for example, data cleaning and further data coding) were carefully documented and verified by another external research firm prior to publication to ensure that no changes to the original data occurred other than those documented in the report.
- A restricted-use data file was prepared and can be used by other researchers to replicate the analyses included in this report or to conduct additional analyses.

The report is organized as follows. Chapter 2 presents the study design and methodology, including the timeline, target population, recruitment, random assignment, sample size, baseline characteristics of the sample, data collection methods, outcome measures, and data analysis methods. Chapter 3 describes the professional development offered to teachers and their implementation of the 6+1 Trait Writing model. Chapter 4 presents the results of the impact analysis. Chapter 5 presents the results of additional exploratory analyses. Chapter 6 summarizes the report's findings and identifies the limitations of the study. The appendixes provide additional details on various topics.

2. Study design and methodology

This chapter describes the research design, timeline, sample recruitment, compensation for activities performed by school personnel, random assignment procedures, baseline equivalence of the control and treatment groups, data collection methods, and outcome measures. It also discusses data analysis methods, missing data, sensitivity analyses, and exploratory analyses.

A multisite cluster randomized trial

A multisite cluster randomized trial was used to assess the effects of 6+1 Trait Writing on the writing performance of a sample of grade 5 students in Oregon. Schools in which grade 5 teachers volunteered to participate in the study were randomly assigned to treatment and control conditions. Teachers in the treatment condition agreed to participate in professional development on the 6+1 Trait Writing model and to integrate this approach into their existing writing curriculum. Use of 6+1 Trait Writing as an instructional enhancement rather than a replacement for existing writing curricula is consistent with the recommendations of the developer. Teachers in the control condition continued to provide their usual writing curriculum and instruction.

The study design employed random assignment of schools rather than random assignment of individual students or teachers, in order to follow the implementation recommendations of the developer. The intervention is intended to be implemented schoolwide whenever possible. Professional development activities include team planning and other collaborative activities.

Two cohorts of schools and teachers participated in the study across two consecutive years. The first cohort was recruited and randomly assigned during 2007. Grade 5 teachers assigned to schools in the treatment condition were offered the 6+1 Trait Writing professional development during summer 2007 and throughout the 2007/08 school year. Data from both treatment and control schools in the first cohort were collected during the fall and spring of the 2007/08 school year.

The same pattern was followed one year later for the second cohort. Schools were recruited and randomly assigned during 2008. Grade 5 teachers assigned to schools in the treatment condition were offered the 6+1 Trait Writing professional development during summer 2008 and during the 2008/09 school year. Data from both treatment and control schools in the second cohort were collected during the 2008/09 school year. Data from both cohorts were combined for the statistical analysis.

Study timeline

A timeline of study activities related to participant recruitment, random assignment, treatment delivery, and data collection is presented in table 2.

Table 2. Timeline of the 6+1 Trait Writing effectiveness study, 2007–09

Cohort/date	Task
Cohort I	
August 2007	Recruitment, agreement to participate
	Random assignment of schools to condition
	Baseline survey data collected from all participating teachers
	Initial professional development for teachers in treatment condition schools (three-day large-group session)
September 2007	Pretest data collected from students in treatment and control schools
October 2007–April 2008	Continued professional development for teachers in treatment condition schools (three one-day large-group sessions, telecommunication support as requested)
January 2008	First follow-up survey data collected from all participating teachers
May 2008	Posttest data collected from students in treatment and control condition schools
	Final follow-up survey data collected from all participating teachers
Cohort II	
Spring 2008	Recruitment, agreement to participate
	Random assignment of schools to condition
August 2008	Baseline survey data collected from all participating teachers
	Initial professional development for teachers in treatment condition schools (three-day large-group session)
September 2008	Pretest data collected from students in treatment and control condition schools
October 2008–April 2009	Continued professional development for teachers in treatment condition schools (three one-day large-group sessions, telecommunication support as requested)
January 2009	First follow-up survey data collected from all participating teachers
May 2009	Posttest data collected from students in treatment and control condition schools
	Final follow-up survey data collected from all participating teachers

Source: Authors.

Target sample size, population, and recruitment methods

A statistical power analysis was conducted in November 2006 to determine the number of schools and students needed to detect a minimum treatment effect size. Details of this analysis are in appendix B. The power analysis was based on a random effects model for schools.

Statistical power refers to the sensitivity of a design to detect treatment effects (Cohen 1988; Schochet 2005). To determine the level of statistical power for the study, researchers estimated the minimum detectable effect size under different scenarios. In collaboration with the Oregon Department of Education, they obtained student writing assessment data with a structure very similar to that proposed for the current study, which they modeled in order to estimate the expected degree of similarity of students within schools and the effect size for a school-level covariate (prior year school-level performance on the state writing assessment). These figures, calculated using the 2005/06 grade 4 Oregon writing assessment data, were subsequently used as input parameters for the power analysis. The minimum detectable effect size was then estimated given varying numbers of schools and varying numbers of students within schools.

The power analysis indicated that a minimum of 54 schools would be needed in order to have adequate statistical power to detect a difference of at least 0.25 standard deviations between treatment and control schools. This minimum detectable effect size was chosen in order to focus on identifying interventions that have a large enough impact to produce substantial changes in student performance. For comparison, the average effect sizes for adolescent writing instruction strategies reported in the *Writing Next* meta-analysis ranged from 0.25 for the study of models of good writing to 0.82 for teaching students strategies for planning, revising, and editing their essays (Graham and Perin 2007a, 2007b). For grade 5 students, typical annual growth in reading skill translates to an effect size of 0.32; typical annual growth in mathematics performance translates to an effect size of 0.41 (Bloom 2007). The study reported here was designed to detect a treatment effect that would improve student writing skills as much as other strategies that have been recommended based on reviews of well-designed empirical studies.

Initial recruitment plans called for a sample of 64 schools, each with a minimum of 30 students in grade 5, in order to accommodate possible attrition. However, during early discussions with Oregon school districts, it became clear that many districts would participate only if all of their schools, including those with fewer than 30 grade 5 students, could be involved in the study. To accommodate the district requests to include small schools in the study while still preserving the desired level of statistical power, researchers included 74 schools in the final sample, each with a minimum of 20 students in grade 5.

The final sample of 39 treatment schools and 35 control schools yielded an expected minimum detectable effect size of approximately 0.23. This means that given the sample size and methods, the study was expected to have an 80 percent chance of detecting a difference between the writing essay scores of the treatment and control groups if this

difference existed and was 0.23 standard deviations or larger. Another way of understanding effect size is in terms of improvement in average percentile scores. For example, an intervention with an effect size of 0.23 would increase average percentile scores from 50 to 59. (For a table of effect sizes and corresponding increases from percentile scores of 50, see www.bestevidence.org/methods/effectsize.htm.)

Oregon was chosen for the study for practical reasons. The Pacific Northwest region has a preponderance of small districts separated by large distances that present logistical challenges for conducting professional development and maintaining oversight of data collection activities. The proximity of Oregon to the study developers and the distribution of schools within the state minimized these challenges.

School eligibility for the study was based on three criteria:

- Trait-based approaches to writing instruction were not currently used and had not been used recently. Participating schools were required to have no recent professional development in 6+1 Trait Writing or similar trait-based writing approaches and no current writing program based on such models. Information from the developers of 6+1 Trait Writing was used to exclude schools in which teachers had been trained within the previous three years. However, as there are other providers of trait-based writing materials and training, members of the research team also interviewed district and school personnel to determine whether any similar professional development had been provided for teachers within the previous three years and whether a trait-based approach was already in use. This information was used to exclude schools in which the 6+1 Trait Writing model or closely related approaches to writing instruction were already in place.
- At least one grade 5 teacher was willing to participate in the research protocol, and the principal was supportive of the school's inclusion in the research.
- At least 20 grade 5 students in the classroom(s) of participating teachers would be available to participate during the data collection phase of the study.

Across both years of the study, 38 Oregon school districts were contacted, including 255 schools with grade 5 classrooms. The sample included 19.4 percent of the 196 Oregon school districts and 33.9 percent of the 752 Oregon schools serving grade 5 students.

Recruitment began with the larger districts in the state, for practical and logistical reasons. In order to facilitate cost-effective group professional development sessions, geographic proximity among participating schools was also considered. For the first cohort of schools (schools in which data were collected during the 2007/08 school year), recruitment focused on schools located in the Willamette Valley area of northwestern Oregon and in several central Oregon districts. For the second cohort of schools (those in which data were collected during the 2008/09 school year), recruitment focused on schools located in the inland and coastal areas of southern Oregon. After some of the state's larger districts agreed to participate, smaller districts that were within a short driving distance were contacted in order to increase the sample size while minimizing the

costs of providing professional development. Teachers from these districts were able to attend professional development sessions without incurring overnight travel expenses.

All schools with grade 5 classrooms were initially considered. Within the 38 districts contacted, 28 schools were disqualified because they were too small to provide a stable sample of 20 or more grade 5 students, and 40 were disqualified because of previous exposure to trait-based writing. Among the schools that met the inclusion criteria, 147 schools (or their district representatives) declined to participate, resulting in an initial recruitment of 75 schools from 22 districts. Entire districts that declined accounted for 112 of the 147 schools that declined (76 percent). Of the schools that declined to participate, 82 (56 percent) were located in three large districts that declined at the district level. One school that initially agreed to participate withdrew after being dissolved and reconstituted with new personnel, reducing the sample size to 74.

In August 2007, after the study design was approved by the U.S. Office of Management and Budget and the Institutional Review Board for the Protection of Human Subjects in Research at Education Northwest, formal invitations were issued to districts and schools that had indicated an interest in participating. Principals of participating schools signed memoranda of understanding that detailed the conditions of participation, including random assignment to treatment or control conditions, required teacher participation in professional development sessions for treatment schools, implementation of the 6+1 Trait Writing approach in schools that were randomly assigned to the treatment condition, and data collection from all participating schools, teachers, and students. Contracts were signed with each participating district detailing the conditions for participation as well as assistance the district would provide during the study and remuneration the district would receive for the staff time involved. A total of 54 schools signed up to participate as part of Cohort I of the study, with professional development and data collection to occur during the 2007/08 school year.

In spring 2008, another 21 schools were recruited to participate in the study during the 2008/09 school year (Cohort II). In summer 2008, after random assignment to conditions, one of the schools assigned to the control condition was dissolved and reconstituted with a new principal and teaching staff; this school subsequently declined to participate in the study. Cohort II proceeded with 20 participating schools during 2008/09. The total number of participating schools during data collection and analysis was thus 74.

Table 3 presents demographic characteristics of the 74 participating schools, the 147 eligible schools that declined to participate, and all Oregon schools. The average proportion of students from racial or ethnic minority groups was 23.7 percent across all study schools; the state mean was 29.6 percent, and the mean for schools that declined was 36.9 percent. The average proportion of students eligible for free or reduced-price lunch in the study sample was 48.9 percent; the state mean was 41.5 percent, and the mean for schools that declined was 53.1 percent. In 2007, the year before the study began, the average proportion of grade 4 students in the sample schools who were at or above proficiency on the Oregon Assessment of Knowledge and Skills was 41.6 percent for writing, 78.6 percent for reading, and 70.8 percent for mathematics. The state means

were 43.8 percent for writing, 79.0 percent for reading, and 71.0 percent for mathematics. The means in schools that declined to participate in the study were 44.7 percent for writing, 77.8 percent for reading, and 70.5 percent for mathematics.

Table 3. Demographic characteristics of participating schools and schools that declined

Characteristic	Sample mean and standard deviation[a]	Schools that declined mean and standard deviation[a]	State mean and standard deviation
Percentage of racial/ethnic minority students[b]	23.7 SD = 15.5 (*n* = 74)	36.9 SD = 23.2 (*n* = 147)	29.6 SD = 19.6 (*n* = 1,273)
Percentage of students eligible for free or reduced-price lunch	48.9 SD = 19.8 (*n* = 74)	53.1 SD = 24.2 (*n* = 147)	41.5 SD = 21.9 (*n* = 1,274)
Percentage of students at or above proficiency on Oregon grade 4 writing assessment in spring 2007	41.6 SD = 14.3 (*n* = 71)	44.7 SD = 18.2 (*n* = 143)	43.8[c] SD = 17.8[c] (*n* for mean = 739; *n* for SD = 707)
Percentage of students at or above proficiency on Oregon grade 4 reading assessment in spring 2007	79.0 SD = 9.2 (*n* = 71)	77.8 SD = 12.0 (*n* = 143)	78.9[c] SD = 10.6[c] (*n* for mean = 745; *n* for SD = 640)
Percentage of students at or above proficiency on Oregon grade 4 mathematics assessment in spring 2007	71.5 SD = 11.6 (*n* = 71)	70.5 SD = 15.1 (*n* = 143)	71.0[c] SD = 13.4[c] (*n* for mean = 743; *n* for SD = 683)

a. Means in the first two rows represent the total number of schools that were included in the study or that were eligible but declined to participate. For the other rows, the number of schools is lower because three sample schools and four declining schools had no grade 4 students.
b. Includes students identified as belonging to one of the following categories: American Indian/Alaska Native, Asian/Pacific Islander, Hispanic, or Black non-Hispanic.
c. Statewide means for all Oregon schools are reported by the Oregon Department of Education. State standard deviations were calculated by the authors based on 2007 Oregon Department of Education data released to the public, which suppresses data from some schools for student privacy protection, usually because more than 95 percent of students met standards.
Source: Authors' analysis of 2007 Oregon Department of Education data.

Incentives to participate in the study

No financial or other incentives were offered or provided to individuals, schools, or districts strictly as rewards for their participation. However, several components of the study design may have been viewed as incentives to participate.

Teachers in both treatment and control group schools were offered professional development in the 6+1 Trait Writing model at no cost (control group teachers were offered the professional development the year after data collection was concluded). Teachers, who were not under contract to their districts during the summer, were compensated at the rate of $150 per day for their attendance at the three-day summer

training session. Districts were reimbursed for the cost of substitute teachers provided in order to release participating teachers for three one-day training sessions during the school year. Principals, literacy specialists, librarians, and special education teachers who worked with grade 5 students were also invited to attend the professional development sessions. All training materials, trainers, and access to telecommunications support during the year were provided at no cost to participants.

In addition, each district was compensated at the rate of $2,500 per school for personnel time and related costs involved in providing assistance with planning and implementing the professional development and handling extensive data management tasks required to protect student confidentiality.

Random assignment, study participants, and attrition

Schools and districts agreed to participate in the study; grade 5 teachers volunteered to participate in the study before random assignment. In 73 of the 75 schools that were initially randomly assigned (including one school that later dropped out), all grade 5 teachers volunteered to participate.

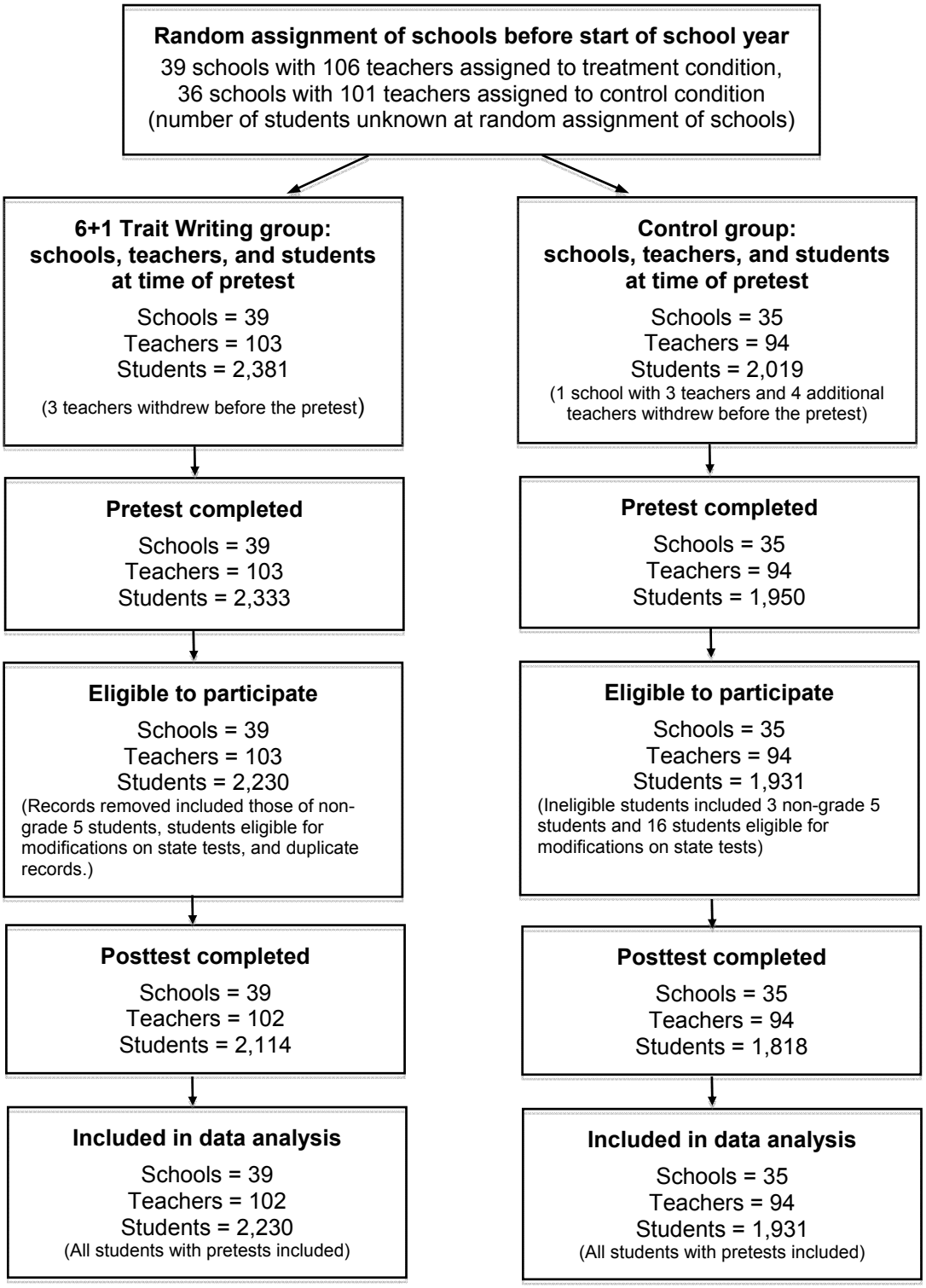
After the participating schools and teachers were identified, random assignment of schools to experimental conditions was done by Chesapeake Research Associates, an independent external research group that had no knowledge of the particular schools involved. Cohort I schools were randomly assigned during summer 2007, before the initial summer training in August 2007 and before initial student data collection, which occurred in September 2007. Cohort II schools were randomly assigned as their districts agreed to participate during spring 2008, before the initial summer training in August 2008 and before initial student data collection, which occurred in September 2008.

Random assignment was done within districts and within pairs of schools based on the proportion of students eligible for free or reduced-price lunch within each school. For example, if a district had six participating schools, the two schools with the highest percentages of students eligible for free or reduced-price lunch were paired, and one of the schools was assigned to the treatment condition and the other to the control condition. This procedure was then followed for the two schools with the next highest proportion of students eligible for free or reduced-price lunch, until all schools within a district had been assigned to either the treatment or control condition. In districts in which an odd number of schools participated, the school with the lowest free or reduced-price lunch rate was randomly assigned to either the treatment condition or the control condition. Each school thus had a 50 percent probability of being assigned to either the treatment or the control condition.

Maintaining the integrity of assignment to conditions and tracking any differential attrition of participants within conditions are key considerations in the conduct of experimental studies. Figure 1 presents the number of participating schools, teachers, and students at each phase of this study, using a flowchart adapted from the Consolidated Standards on Reporting Trials (CONSORT) statement. For this study, all students present

in participating classrooms were given the pretest as a classroom activity; eligibility for the study was determined later, as described below.

Figure 1. Sample size at various phases of the study



Source: Table format adapted from the Consolidated Standards on Reporting Trials (CONSORT) statement (www.consortstatement.org).

Random assignment phase. Thirty-nine schools (including 106 teachers) were randomly assigned to the treatment condition, and 36 schools (including 101 teachers) were randomly assigned to the control condition. At the time of random assignment, the number of students that would enroll for the following school year was not known.

In July 2008, after random assignment, one of the Cohort II control schools (with three teachers scheduled for participation) was dissolved and reconstituted with a new principal and teaching staff. The school then declined to participate in the study; the three teachers scheduled for participation were no longer employed and withdrew from the study. The total sample was thus reduced to 39 schools and 106 teachers in the treatment condition and 35 schools and 98 teachers in the control condition. For each cohort, randomization of schools and initial professional development activities for treatment group teachers occurred during the summer, when the exact number and the characteristics of students who would attend these schools during the year of the study were still unknown.

Four control group teachers and three treatment group teachers who had volunteered for the study moved to different grades or schools by the time of the pretest and withdrew from the study, reducing the total number of teachers to 103 in the treatment condition and 94 in the control condition at the time of the baseline student assessments. (The new grade 5 teachers in these schools continued to participate in the study, so the number of schools remained unchanged.)

After the beginning of the school year, classroom rosters were completed to obtain the number of students in the classrooms of participating teachers who were available to participate in the study. The rosters included students who were receiving writing instruction in the regular classroom; they did not list students who were being pulled out of the classroom during writing instruction because they were English language learner students or special education students requiring specialized instruction in separate classrooms. The rosters included 2,381 students in treatment group schools and 2,019 students in control group schools.

Pretest phase. All schools originally assigned to conditions (except the one control condition school that withdrew before the pretest) continued to participate in the study from the pretest through the posttest. Schools forwarded pretests directly to Chesapeake Research Associates, which relabeled the student essays with coded identification numbers and sent them to the essay rating team at Education Northwest. Students who did not complete the pretest were not included in the study.

Of the 4,400 students who were receiving writing instruction in study classrooms at the time of the pretest, 117 were absent at the time the pretest was administered and were not included in the study. A total of 4,283 students completed the pretest; 2,333 of them were in treatment condition schools and 1,950 were in control condition schools.

Because of the characteristics of the study (evaluation of classroom instructional methods using assessment procedures typical for the school setting) and the fact that personally

identifiable information about individual students was never transferred from the districts to the researchers, parental consent and student assent were not required.

Not all of the 4,283 students who completed the pretest received a pretest score. Students lacking a pretest score included 169 students who returned an essay that could not be scored because it was blank, illegible, written in a language other than English, or deemed by the scorers as "too short to score." Before data analysis, these pretests were assigned the lowest possible score.

Student eligibility to participate in the study. Student eligibility to participate in the study was established retrospectively, using the demographic information provided by schools and compiled by Chesapeake Research Associates. Among the 4,283 students who completed the pretest, 99 students in multigrade classrooms were not in grade 5; these students (96 from the treatment condition and 3 from the control condition) were removed from the study data. Another 22 students (6 from the treatment condition and 16 from the control condition) were eliminated from the study because they were eligible for "modifications" on the Oregon statewide assessment (that is, the severity of their disability prevented them from taking the same state tests as other students). Students with disabilities who were eligible for "accommodations" on statewide assessment (that is, students who take the same test as other students but are allowed minor alterations in test-taking procedures) were included in the study. These eligibility criteria were established before any data collection occurred.

One (treatment group) student record was removed because it was found to be a duplicate record. This resulted in a total of 4,161 students who were eligible for the study.

Posttest phase. All schools and classrooms that participated in the pretest also participated in the posttest. Among the 2,230 students in treatment condition schools who completed the pretest and were eligible to participate in the study, 2,114 (94.8 percent) completed the posttest. Among the 1,931 students in control condition schools who completed the pretest and were eligible to participate in the study, 1,818 (94.1 percent) completed the posttest.

During the study, 14 students transferred from treatment schools to control schools, and 9 students transferred from control schools to treatment schools. In all of these cases, the students' pretest and posttest scores were analyzed as part of the school to which they had been originally assigned. No schools or teachers moved from one treatment condition to the other during the study.

Data analysis phase. The sample included in the final data analysis consisted of students from all of the originally participating schools and classrooms (except for the one control condition school that withdrew from the study before the beginning of professional development and before the pretest). All participating schools continued as part of the experimental condition to which they had been randomly assigned.

The final sample included 2,230 students in treatment condition schools and 1,931 students in the control schools. All eligible-to-participate students who completed the pretest and the posttest were included in the analysis. In addition, for students who did not complete a posttest but did complete a pretest, their posttest scores were statistically imputed based on other data.

Attrition rates. After random assignment and before any study-related activities began, one school dropped out of the study after it was reconstituted and the staff replaced. The overall school attrition rate was thus 1.3 percent (zero for the treatment group, 2.8 percent for the control group).

After random assignment, 10 teachers dropped out of the study, including 3 in the control group school that was reconstituted and 7 others (3 treatment group and 4 control group teachers). The overall teacher attrition rate was thus 4.8 percent (2.8 percent for the treatment group, 6.9 percent for the control group).

For the purpose of this study, the student-level attrition rate was defined as the number of eligible-to-participate students who did not complete the posttest divided by the number of eligible-to-participate students who completed the pretest. The overall student attrition rate was 5.5 percent (5.2 percent for the treatment condition, 5.9 percent for the control condition).

Baseline equivalence of treatment and control groups

The purpose of random assignment is to create treatment and control groups that are equivalent at the outset of a study, so that any differences observed later can be attributed to the effect of the treatment. In practice, random assignment can sometimes result in treatment and control groups that differ by chance at the beginning of the study.

In order to determine whether the random assignment of schools to a condition resulted in groups that were similar at the beginning of the study, researchers compared the baseline measures of school, teacher, and student characteristics (table 4). Baseline data met the standard statistical assumptions for these tests (data were normally distributed, with equal variances and no influential outliers). The 74 schools in this baseline sample were also included in the final analytic sample.

At baseline the treatment and control group schools had similar proportions of students eligible for free or reduced-price lunch, as well as similar proportions of students from racial or ethnic minority groups. They did not differ significantly on the proportion of girls attending the school or the proportion of students who scored at or above proficiency on the 2007 grade 4 Oregon assessments in reading, mathematics, or writing. There was no difference between the groups on the study-administered pretest of student writing proficiency.

There were some statistically significant differences between treatment and control group schools at the beginning of the study. Teachers in control group schools reported having an average of 14.5 years of teaching experience compared to 10.8 years of teaching experience for those in treatment group schools. In addition, teachers in control group schools reported having an average of 12.9 years of experience specifically teaching writing, compared to 10.0 years of experience teaching writing for those in treatment group schools. Control group teachers also reported that their students spent less time practicing writing in class (3.4 hours versus 4.2 hours). These findings were reported on teacher surveys administered during the summer before the year in which the intervention occurred. The presence of these baseline differences between treatment and control teachers could potentially confound the impact estimate. Therefore, three covariates were added to the planned benchmark analysis model in order to adjust the impact estimate for these exogenous differences between schools. The originally planned benchmark model, without these covariates, was also analyzed; the results of both analyses are presented in chapter 4.

Table 4. Baseline characteristics of study schools, teachers, and students

Characteristic	Treatment group	Control group	Difference	Test statistic
School proportion of racial/ethnic minority students[a] *(percent)*				
Mean	24.1	25.4	–1.3	$t = –0.34$ $p = .78$
Standard deviation	16.22	16.24		
Sample size	39	35		
School proportion of students eligible for free or reduced-price lunch (percent)				
Mean	48.6	49.3	–0.7	$t = –0.16$ $p = .87$
Standard deviation	19.29	20.72		
Sample size	39	35		
School proportion of girls (percent)				
Mean	47.2	47.6	–0.4	$t = –0.33$ $p = .74$
Standard deviation	0.06	0.04		
Sample size	39	35		
School proficiency rate on Oregon grade 4 writing assessment in spring 2007[b]				
Mean	39.1	40.5	–1.9	$t = –0.43$ $p = .37$
Standard deviation	13.64	14.10		

Characteristic	Treatment group	Control group	Difference	Test statistic
Sample size	38	33		
School proficiency rate on Oregon grade 4 reading assessment in spring 2007[b]				
Mean	79.6	80.0	–0.4	$t = -0.18$ $p = .93$
Standard deviation	9.36	10.00		
Sample size	38	33		
School proficiency rate on Oregon grade 4 mathematics assessment in spring 2007[b]				
Mean	73.4	69.9	3.6	$t = 1.26$ $p = .33$
Standard deviation	9.94	13.75		
Sample size	38	33		
Student scores on study-administered pretest of writing proficiency[c]				
Mean	3.59	3.67	–0.08	$t = -1.56$ $p = .12$
Standard deviation	0.76	0.78		
Sample size	2,230	1,929		
Teacher years of teaching experience[d]				
Mean	10.8	14.5	–3.7	$t = -2.44$ $p = .02$
Standard deviation	8.37	10.07		
Sample size	94	90		
Teacher years of experience teaching writing[d]				
Mean	10.0	12.9	–2.9	$t = -2.16$ $p = .03$
Standard deviation	7.86	9.19		
Sample size	94	89		
Teacher-reported weekly in-class hours students spend practicing writing[d]				
Mean	4.2	3.4	0.8	$t = 2.08$ $p = .04$
Standard deviation	3.06	1.55		
Sample size	89	87		
Teacher-reported weekly hours students spend completing				

Characteristic	Treatment group	Control group	Difference	Test statistic
homework that involves significant writing[d]				
Mean	1.0	1.0	–0.01	$t = –0.12$ $p = .90$
Standard deviation	0.93	0.81		
Sample size	87	82		

a. Percentage includes students identified as belonging to one of the following categories: American Indian/Alaska Native, Asian/Pacific Islander, Hispanic, or Black non-Hispanic.
b. Means for proficiency rates on grade 4 Oregon statewide assessments include only 71 schools, because 3 schools were middle schools and had no grade 4.
c. Based on the preimputation dataset; *t*-test calculated accounting for clustering.
d. Teacher reports are from baseline surveys completed by 96 of the 103 treatment group teachers (93.2 percent) and 92 of the 94 control group teachers (97.9 percent). Sample size for some items is lower because of item nonresponse; *t*-test calculated accounting for clustering.
Source: Authors' analysis, based on data described in text and Oregon Department of Education 2007 statewide assessment data.

Data collection instruments and procedures

Data collection instruments included student rosters, a student essay booklet, teacher instructions for administering the student essay assessment, and teacher surveys. These instruments are briefly described below, along with information about how they were used to collect data for the study (see table 2 for the data collection schedule).

Classroom rosters. Classroom rosters in the form of both electronic spreadsheets and printed paper copies were provided to each school and used to track information about students who completed pretest and/or posttest essays. Each participating teacher received a classroom roster that included a series of prerecorded unique student identification numbers created for the study. In addition to the participating classroom teachers, districts assigned a "site coordinator" to each school to assist with data collection and data management, including completion of the student rosters.

The classroom roster included a series of columns for recording the name, grade, ethnicity, and eligibility for accommodations or modifications on assessments (because of special education or English language learner issues). Also included were columns for recording whether the student completed the pretest and posttest and, if applicable, the reason why students did not complete or take the test. The identification numbers were structured so that the research team could determine each participating student's district, school, and classroom teacher. Researchers did not have access to student names.

For both the pretest and the posttest, student essay booklets were supplied to the schools with these unique identification numbers printed on them. Teachers and site coordinators were responsible for ensuring that each student completed the pretest and posttest using the student essay booklet with the printed identification number assigned to him or her on the classroom roster. This coding system allowed researchers to link the student

information recorded on the classroom roster with each student's pretest and posttest data.

Throughout the year, two copies of each roster were maintained at the school for safekeeping, one kept by the teacher or site coordinator, and the other kept by the principal. At the end of the data collection period, a copy was made for the research team, with the column of student names removed. In this way, the student demographic information and other data were transmitted to the research team in a way that facilitated the linkage of these data to the pretest and posttest essays. Student names or other information that could allow the identification of particular students were never transmitted to the researchers.

Student essays. For both the pretest and the posttest, participating students were provided with a paper booklet containing a prompt, to which they were asked to respond by writing an essay. The booklet (a copy appears in appendix C) included pages for planning the essay, creating a first draft, revising this draft, and creating a final version. These activities were to take place over three class periods of about 45 minutes each, on three consecutive or near-consecutive school days.

The prompt used for the study was the same for both the pretest and the posttest: *Think of a skill you have learned that has made your life easier or more fun. Write a letter telling about your skill, explain how you learned it, and why you think it is important"* This prompt was created for the study, based on the RAFT model of writing prompts (Santa 1988; see appendix A, box A1, item 7). Use of the same prompt throughout the study was intended to avoid variation in student performance that could be related to the characteristics of different writing modes (for example, explanatory writing versus persuasive writing) or specific prompts. Rather than narrowly specifying what must be explained in the essays, students were allowed to choose what they wished to explain, within the general guideline provided by the prompt. Along with the essay prompt and the booklet for organizing their work, students received verbal instructions from their teacher to guide them through a three-day process of planning, drafting, and completing an essay. Teachers were instructed not to give any assistance to students beyond the information provided in the verbal instructions.

Schools were provided with mailing envelopes and instructions on how to send the student essays to the research team. At the conclusion of the pretest, and again at the conclusion of the posttest, each site coordinator sent the student essays directly to Chesapeake Research Associates, which created a separate coding system that was used to translate the unique identification numbers of students into a separate unique identification numbering system. Each student essay was photocopied, labeled with the new unique identification number, and shipped to REL Northwest, where the essays were scored. Pretest and posttest essays were mixed together for this phase of the study, and essays from treatment and control condition schools were mixed together. All student essays from each cohort (pretests and posttests from control and treatment schools) were delivered to the rating team at the same time.

This process was used to ensure that raters—and the entire REL Northwest research team—were blind to the source of the essays during the scoring process. Each student essay was scored by two teams of raters, to produce the scores that were used in the data analysis. The raters did not know whether a particular essay was a pretest or a posttest or whether it came from a treatment or a control condition school. (More information on the rating process is included in the section on outcome measures, later in this chapter.) After all scoring was completed, the original REL Northwest unique identifiers were restored so that data analysis could proceed.

Pretest writing samples were collected in September of each data collection year; posttest writing samples were collected the following May. Specific directions for test administration were provided to teachers, explaining the purpose of and procedures for each of the three days. Teachers were also provided current copies of both the accommodations table and the modifications table Oregon used for writing test administration. Teachers were asked to record whether students were eligible for accommodations or modifications based on the same procedures used for the Oregon state writing assessment.

Teacher instructions for proctoring the student essay writing sessions. The student essay writing sessions were proctored by participating teachers in both control and treatment groups, in a manner similar to that used for administering the Oregon statewide writing assessment in grades 4, 7, and 10. Students were given three class periods on three different days to work on their essays, to provide the opportunity for a natural writing process, including planning, drafting, and revision.

The test administration directions gave explicit instructions for how each of the three class periods on each of the three assessment days should be structured and how the process should be explained to students (see appendix C). The use of dictionaries and thesauruses was allowed; the use of spelling or grammar checkers was not. Students were to work on their own, without assistance from peers or teachers. All student essay booklets were kept by the teacher between class sessions; students were not allowed to work on them at home or during other parts of the school day.

Although the use of classroom teachers as proctors for the assessment could theoretically allow teachers to bias the results of the study—and readers should bear this in mind—several considerations may limit the likelihood of such bias. Teachers also proctor the Oregon state assessments, using a process very similar to that used in this study; as part of their participation in the statewide assessments, they sign a Test Administrator Assurance of Test Security, which addresses the integrity of the test-proctoring process. Each teacher was provided with explicit directions, specific rules for what kinds of assistance could and could not be provided for students, and verbatim scripts to follow in administering the assessments used for the study (see appendix C). Although some teachers may have failed to implement these procedures properly, in the absence of evidence to the contrary, it is reasonable to assume that deviations were randomly distributed across the control and treatment schools, resulting in no overall bias in the impact estimate.

Teacher survey. A teacher survey was used to gather information about teacher classroom practices related to writing instruction, as well as the number of years teachers had been teaching in general and the number of years they had been teaching writing. A copy of the teacher survey is in appendix D.

The survey, which was created for this study, was designed to measure the degree to which teachers implemented the approach to writing instruction that was the focus of the study. Each survey item asked teachers about an aspect of 1 of the 10 instructional practices that are key elements in the 6+1 Trait Writing intervention. However, the items were worded so that they could be answered by teachers with only a basic awareness of the general idea of trait-based writing, as would be gained from familiarity with the Oregon statewide standards and assessment system. The questions did not address details of instructional practices specific to the 6+1 Trait Writing intervention. Early versions of the survey were pilot tested with small groups of teachers to ensure that the items made sense to teachers regardless of their prior knowledge of the intervention.

The survey was intended to provide a measure of the extent to which the practices promoted in the intervention were actually implemented by teachers in the treatment condition group and the extent to which these practices were implemented by the control group teachers despite their lack of experience with the 6+1 Trait Writing model.

Teachers completed the survey at three points in time: before the school year in which student data were collected (during which teachers in the treatment group received professional development), at midyear, and at the end of the year. The baseline survey was administered to teachers before the study-sponsored professional development was provided. Paper copies of the teacher survey and prepaid return envelopes were delivered to the site coordinator for each school, which distributed them to participating teachers. Teachers completed the surveys independently and mailed them directly to the researchers.

At baseline, 93.2 percent of teachers in the treatment condition and 97.9 percent of teachers in the control condition completed and returned the surveys. At midyear, 95.1 percent of teachers in the treatment condition and 94.7 percent of teachers in the control condition completed and returned the surveys. At the end of the year, 88.2 percent of teachers in the treatment condition and 96.8 percent of teachers in the control condition completed and returned the surveys. These figures exclude three copies of the baseline surveys and one copy of the end-of-year survey that were returned without identifying information and could not be classified by experimental group.

Outcome measures

Data collected during the study included student essays and teacher surveys.

Student essays. The outcome measure for the primary (confirmatory) research question was a single score for overall writing quality that was produced by a team of raters, using the process described below. This score was also used for the second of two exploratory analyses, which examined gender and ethnicity differences. A team of raters also produced separate ratings for each of six traits of writing, which were used in the first exploratory analysis.

Education Northwest maintains a pool of writing assessment raters who scored the essays without any knowledge of students' experimental condition or whether an essay was a pretest or posttest. This rating team has scored papers for school districts in the United States and foreign countries for more than two decades; it has also scored writing samples provided by job applicants for major corporations. The team is experienced in using trait-based rubrics as well as holistic and client-specific rubrics on persuasive, expository, and narrative modes of writing.

For this study, each student essay was scored separately by two sets of raters, who followed the standard methodology and data management process they use for all writing assessment projects, with the exception of methods described above to keep the raters blind to the experimental condition and the pretest or posttest status of each essay. One scoring team applied the holistic rubric; a second team applied the six analytic rubrics. These two types of rubrics are described below and presented in appendix E.

The primary outcome measure was a slightly modified version of a scoring guide that is used for scoring college entrance examinations; it was not designed to be closely aligned to the intervention program elements. Two raters from the first team conducted independent reviews of each student essay using a holistic rubric (scoring guide or rating scale) to give each essay a single score. This rubric was derived from the scoring guide used by the College Board in 2005 for scoring college entrance examination writing samples on the SAT. (A slightly updated version of this rubric, used in 2010, is available on the College Board website, at http://professionals.collegeboard.com/testing/sat-reasoning/scores/essay/guide.) One phrase within one item on the scoring guide was altered for this study because it was oriented specifically toward scoring a persuasive rather than an expository essay. Student scores resulting from application of the holistic rubric were used as the primary outcome variable for the study.

Each student essay was also rated by two additional raters, using six separate rubrics, one for each of the six core traits in the 6+1 Trait Writing model. These additional outcome measures were included in case their alignment with the intervention might result in more sensitivity to program effects. (The trait of presentation was not scored because students did not have the opportunity to address presentation issues during the assessment—the test booklet required that they use a standard presentation format.) All six trait scales were positively correlated with the holistic scale. In the analysis sample, the bivariate

correlation (Pearson *r*) between the six trait scales and the holistic scale ranged from 0.49 to 0.65 for the pretest and from 0.55 to 0.66 for the posttest. Details of these correlations are provided in appendix F. Combined, the six trait scales accounted for 75.2 percent of the variance in the holistic scale score at pretest and 80.6 percent of the variance in the holistic scale score at posttest. These trait-specific scores were used in the exploratory analysis. Essays not in English and essays of two sentences or less were not scored by the team. They were coded as "unable to rate" and assigned the lowest possible score during the analysis.

For both the holistic and the individual trait scores, each student essay was first scored independently by two raters, who entered their scores in a database. Each rater assigned the essay a score ranging from 1 (indicating a severely flawed essay demonstrating very little or no mastery) to 6 (indicating an essay demonstrating clear and consistent mastery of writing). The final score for the essay was the average of the two, scored in half-point increments. If the scores of the two raters were identical or only one point apart, no further action was taken. If the scores of the two raters were more than one point apart, the essay was flagged and scored by the team leader, who then asked the two original raters to read the essay again and revise their scores in consultation with her. This procedure was used for continual quality improvement of the rating team and as a quality assurance procedure for the individual essay scores. It is a common practice to improve the reliability of ratings of open-ended assessments (Johnson et al. 2005). Details related to the number of essays that could not be scored and interrater reliability are included in appendix G.

Teacher survey. The teacher survey created for this study by the research team included 10 scales corresponding to the 10 classroom practices (also called teaching strategies) emphasized by the intervention. The instrument is reproduced in appendix D. Each scale contained three to seven items. For each item, teachers used a seven-point scale to rate the extent to which they implemented a classroom practice or strategy. Scale scores were created for each of the 10 scales by taking the mean score for all items within the scale for each teacher. At the initial survey administration, internal consistency of the scales ranged from 0.78 to 0.92. Specific values for each administration of each scale are reported in the Fidelity of implementation section in chapter 3 of this report, along with a discussion of the potential problems with the validity and interpretation of these findings.

The survey also included questions about the number of years teachers had been teaching, the number of years they had been teaching writing, the number of hours each week their students practiced writing in class, and the number of hours each week their students spent on homework that included writing. Teachers entered numbers to answer these questions. The data were then averaged across the group. Teachers were also asked to list the writing program they used with their students and any professional development related to the teaching of writing they had experienced in the previous two years.

Data analysis methods

A multilevel statistical analysis was performed to estimate the treatment impact and to answer the confirmatory research question. Student data were analyzed as part of the experimental group to which they were originally assigned. The analysis involved:

- Estimation of the treatment impact as a covariate-adjusted difference between the treatment and the control schools.
- Use of a multilevel model to reflect the nesting of students within schools.
- Sensitivity analyses to determine whether changes in the statistical model would alter the findings.

The multilevel model included two random effects, the school random effect and the student residual. Including the school random effect in the model adjusted the standard error, which was important for calculating an accurate significance test for the treatment effect.

The data came from students attending 74 schools. A paired-randomization method was used to assign schools to the treatment or control condition. Within each participating district, pairs of schools were formed based on their similarity in the proportion of students within each school who were eligible for free or reduced-price lunch (*FRL%*). Proper modeling of the paired randomization—a special case of block randomization—is challenging, because each pair contains only two replications, one in the treatment arm and the other in the control arm. Two modeling approaches were considered: 1) use of pair indicator variables, or 2) use of district indicator variables along with FRL%. The first approach was chosen because it is the most direct way to account for the paired randomization in the experimental design and to control between-pair variance.

The data file for analysis was constructed by Chesapeake Research Associates using writing scores provided by raters who were blind to the experimental condition and to the pretest or posttest status of each student essay. Student essay pretest and posttest writing scores were matched with individual demographic data recorded on the classroom rosters. Chesapeake Research Associates then transmitted the complete raw data file to the research team, which at that time was also provided with information about which school, classroom, student, and survey administration was associated with each particular essay.

During initial data cleaning, the accuracy of input was assessed by examining out-of-range values, plausible means and standard deviations, and outliers. Students not eligible for the study (for example, non-grade 5 students inadvertently included in the raw data) were removed. The extent and distribution of missing data were determined. Because the number of cases with missing data on the posttest was more than 5 percent of total cases (5.2 percent for the treatment group and 5.9 percent for the control group), multiple imputation (MI) was used before the impact analyses to impute both covariates and the posttest outcome variable. The decision to use multiple imputation if more than 5 percent

of cases had missing values had been specified in the analysis plan for the study prior to collecting data.

To ensure that no errors occurred during the cleaning of the raw data and the preparation of the final analytic sample, Chesapeake Research Associates transmitted the original raw data file to Empirical Education, an independent research firm. The study authors transmitted to Empirical Education the criteria and process used for cleaning the data file and preparing the sample for analysis. Empirical Education independently applied these data cleaning procedures to the raw data file and compared the resulting sample size, descriptive means, and standard deviations with those used in the analysis, in order to verify the integrity of the analytic sample. No errors were found.

The multilevel model. This section describes the multilevel model used for the impact analysis. The treatment effect was estimated using a student-nested-within-school multilevel model. The treatment effect was defined as the covariate-adjusted mean difference between the treatment and control conditions. For ease of interpretation, the multilevel model is expressed as a hierarchical linear model (HLM).

Recall that the pairing of schools was done within each district, using FRL% as the pairing variable. In districts with an even number of schools, this procedure resulted in the pairing of all schools. In districts with an odd number of schools this procedure resulted in one "singleton" school that was not part of a pair. In the entire sample of 74 schools there were 14 of these singleton schools.

In order to properly model the experimental design, the analysis model must account for both the pairing of schools and the presence of singleton schools. All the analyses—impact, sensitivity, and exploratory—followed a two-step process. First, the estimate of effect was done separately for the paired schools (n = 60) and for the singleton schools (n = 14). A model reflecting the pairing of schools was used for the paired schools. A simplified version of the model lacking the variables used to identify the pairing was used for the singleton schools. Since singleton schools varied substantially in FRL%, this variable was included in the model used for the singleton schools, in order to improve its statistical precision. Once an estimate of the effect was obtained from paired schools and from singletons, the two estimates were combined using the method of inverse-variance weighting.

Because the majority of the schools were paired, the descriptions of the analysis models that follow refer to the version of the model used for data from the paired schools. The reader may construct the simplified version of the model for the singleton schools by replacing the variables expressing the pairing with FRL%.

Level 1 model (student-level model)

$y_{ij} = \beta_{0j} + \beta_{1j} PRE_{ij} + e_{ij}$

Level 2 model (school-level model)

$\beta_{0j} = \gamma_{00} + \gamma_{01} TRT_j + \gamma_{02} WkWrite_j + \gamma_{03} YrTeach_j + \gamma_{04} YrWrite_j + d(k)*Pair(k) + u_{0j}$

$\beta_{1j} = \gamma_{10}.$

This hierarchical linear model can also be expressed as a linear mixed model. Researchers relied on the following linear mixed model expression to perform the statistical analyses using Stata 11:

$y_{ij} = \gamma_{00} + \gamma_{01} TRT_j + \gamma_{02} WkWrite_j + \gamma_{03} YrTeach_j + \gamma_{04} YrWrite_j + \gamma_{10} PRE_{ij} + d(k)*Pair(k) + u_{0j} + e_{ij}.$

In the following description, the multilevel model is characterized according to its expression in HLM, as this nomenclature may be more familiar to some readers.

The level 1 model specifies the outcome of student *i* in school *j*. Student *i*'s posttest score (y_{ij}) is a linear function of school *j*'s intercept (β_{0j}) and the student's pretest score *(PRE_{ij})*, plus the residual associated with the student (e_{ij}). The pretest score is grand-mean centered; the school intercept is adjusted for the pretest scores of its students. The school intercept reflects the expected posttest score of the student at school *j* whose pretest score was at the grand mean. The residual (e_{ij}) is assumed to be normally distributed with a mean of zero.

The level 2 model specifies school *j*'s intercept (β_{0j}), as well as its slope for the student-level covariate (β_{1j}). School *j*'s intercept is a linear function of the population intercept (γ_{00}) for the posttest score, the treatment effect (γ_{01}), the pair fixed effect (Pair(k)), plus a random effect associated with the school (u_{0j}). In order to adjust for exogenous differences found between treatment and control teachers at baseline, the following covariates from the baseline teacher survey were added: the school average for the weekly teacher-reported hours students spend in class practicing writing ($\gamma_{02} WkWrite_{j)}$, the school average for teacher years of teaching experience ($\gamma_{03} YrTeach_{j)}$, and the school average for teacher years of experience teaching writing ($\gamma_{04} YrWrite_{j)}$. School *j*'s slope for the student-level covariate *(PRE_{ij})* was constrained to be the same across schools, for the purpose of model parsimony. The random effect of school (u_{0j}) is assumed to be normally distributed with a mean of zero.

The primary interest was the estimation of γ_{01}, the treatment effect adjusted for the baseline performance level of the students, exogenous differences found between treatment and control schools, and the pair fixed effect. This represents the estimate of

the treatment effect for the population of schools. A statistically significant treatment effect would suggest that the 6+1 Trait Writing intervention influences student performance in writing. Stata 11's XTREG command was used, along with the option MLE, which requested XTREG to use the maximum-likelihood estimate instead of its default generalized least squares estimate. The estimated treatment effect in a standardized form was also calculated, using the control group standard deviation of the posttest score. This value represents the effect size in Glass's delta.

Sensitivity analysis. Two sensitivity analyses were conducted to determine whether the results of the benchmark analysis would be sensitive to alternative analytic methods. First, the impact analysis was repeated after deleting data from students who did not complete a posttest, rather than imputing their posttest scores. With the exception of this sensitivity test, all other analyses in this report were performed on the full dataset including imputed values.

For the second sensitivity analysis, the three covariates that had been used to adjust for baseline differences in the school averages for teacher years of teaching experience, teacher years of experience teaching writing, and teaching-reported hours students spend in class practicing writing were removed from the model.

Missing data. Before the impact analysis was conducted, a preliminary analysis was performed to determine the extent and patterning of missing data. Multiple imputation procedures were then used to estimate values for missing data points. The details of this process are provided in appendix H. The impact analysis was conducted using the full dataset, including imputed missing values.

Exploratory analyses. Exploratory analyses were performed to answer the following questions:

1. What is the impact of 6+1 Trait Writing on grade 5 student achievement in particular traits of writing?
2. Does the impact of 6+1 Trait Writing on grade 5 student achievement vary according to student gender or ethnicity?

The first question was answered by simply replacing the holistic score in the impact analysis model with each of the six trait scale scores. Because each of the six scale scores was intended to represent a distinct trait, no statistical adjustment to account for multiple comparisons was made; the six exploratory analyses were treated as tests of six theoretically independent hypotheses.

The second question was answered using the holistic scale score as the outcome with a model that included student gender (or ethnicity) as a moderator variable. The moderating effect of student gender or ethnicity was represented by the treatment-by-gender or treatment-by-ethnicity interaction. The main effect of gender or ethnicity was also included in the model.

The student sample consisted of White non-Hispanics (3,163 of 4,161 students, or 76.0 percent of the sample); Hispanics (516 of 4,161 students, or 12.4 percent of the sample); and other students (482 of 4,161 students, or 11.6 percent of the sample). No other ethnic groups exceeded 4 percent of the sample. Consequently, the student sample was recoded into ethnic subgroups in two ways: White non-Hispanics versus all others, and White non-Hispanics versus Hispanics.

The model for each of the subgroup analyses was:

$$Y_{ij} = \gamma_{00} + \gamma_{01}TRT_j + \gamma_{02}WkWrite_j + \gamma_{03}YrTeach_j + \gamma_{04}YrWrite_j + \gamma_{10}PRE_{ij} + \gamma_{20}\mathrm{SUB2}_{ij} + \gamma_{21}TRT_j\,SUB2_{ij} + \mathrm{d(k)*Pair(k)} + u_{0j} + e_{ij.}$$

The moderator effect is expressed as a cross-level interaction between the treatment dummy *(TRT)* and the dummy for the nonreferent subgroup *(SUB2,* when *SUB1* is the referent).

3. Implementing the 6+1 Trait Writing intervention

This chapter describes the professional development provided to teachers as part of the treatment and presents the results of teacher surveys used to measure the extent to which teachers reported using the instructional strategies in their classrooms. Professional development was provided by the model developers. Beyond the normal technical assistance offered (email and telephone consultation, web-based examples and materials), the developers provided no oversight or treatment, allowing teachers and their administrators to determine how the model was implemented in the classroom.

The 6+1 Trait Writing professional development format

Teachers in the treatment group attended a three-day summer institute that provided comprehensive training, planning time, and resource materials. During the school year, participants attended three additional one-day workshops to further their understanding of the approach and to plan trait-based activities. Learning objectives and detailed agendas for these institutes and workshops are in appendix A.

Teachers also had access to online support resources throughout the year to help them integrate the model into their existing classroom writing instruction. These resources included sample student papers and opportunities to practice scoring student writing, including comparing their scores to those of expert raters; suggested reading material to illustrate particular traits of writing; and access to email correspondence and telephone conversations with 6+1 Trait Writing trainers.

The three-day summer institute was used to introduce the model of writing instruction and assessment to teachers. Training included a brief history of the model, a review of rubrics in general, and a review of the specific rubric used for statewide testing in Oregon. Student papers were scored, and a recommended cycle of instruction was shared, along with planning charts designed to help teachers plan their instruction for the first three months of the school year. The cycle of instruction, which teachers were asked to use with their students throughout the year, provided the following instructional sequence for delivering lessons focused on a particular trait:

- Using the rubric to plan.
- Teaching the language by rewriting the rubric in student-friendly language.
- Scoring papers, justifying the scores by using the rubric, and discussing the scoring process.
- Modeling several trait-based activities or focus lessons.
- Creating a writing prompt.
- Gathering a written product for assessment purposes.
- Measuring overall student improvement as well as improvement in the specific trait.

All 6+1 Trait Writing traits were introduced at the summer institute, and teachers were asked to begin working with their students on all the traits at the beginning of the school year. Teachers were encouraged to emphasize two traits—ideas and organization—over the following three months. The trait of conventions was also emphasized at the summer institute, at each subsequent one-day workshop, and in student instruction throughout the year. This trait was emphasized for two main reasons. First, the conventions of writing include capitalization, punctuation, spelling, and grammar; the 6+1 Trait Writing trainers believe that it is helpful for students and teachers to spread these lessons over an entire school year rather than trying to cover the material in a shorter time period. Second, working on conventions throughout the school year is intended to address the concerns of some teachers, parents, and administrators that correct use of conventions is such an important expectation for writing that conventions should be one aspect of writing instruction at all times.

Ten trait-based classroom writing instruction strategies were introduced, reviewed, and used to model activities and to focus lessons throughout the three-day summer institute:

1. Teaching the language of rubrics for writing assessment.
2. Reading and scoring papers, justifying the scores, and having students do so themselves.
3. Teaching focused revision strategies.
4. Modeling participation in the writing process.
5. Having students read and analyze materials that demonstrate varying writing quality.
6. Giving students writing assignments to respond to effective prompts (that is, prompts that engage students and provide adequate structure, guidance, and context to elicit detailed responses).
7. Weaving writing lessons into other subjects.
8. Teaching students to set goals and monitor their progress.
9. Integrating learning goals for writing into curriculum planning.
10. Teaching ways to structure nonfiction writing.

During the summer institute, hundreds of picture books that teachers could use to teach particular writing traits to grade 5 students were made available for teacher review. Techniques for using picture books in classroom instruction were modeled as each trait was introduced. A particular method of creating prompts to which students would respond in their writing was introduced, and teachers practiced generating their own writing prompts for students. Teachers worked in teams to plan how they would integrate these activities into their instruction over the next three months.

In November, about three and a half months after the initial three-day summer institute, the first of three one-day support workshops was held. This workshop—and the two held

later in the school year—included a troubleshooting session during which teachers discussed their perceptions of how their attempts to use the 6+1 Trait Writing model had and had not been successful during the previous months; teachers also shared their ideas about how to improve their practice. The trainers provided a quick review of the 6+1 Trait Writing model of instruction and assessment and asked teachers to score student papers in order to review the traits that had previously been covered and to provide insight into traits that would be the focus of the new workshop and the next period of instruction. The suggested cycle of instruction for introducing each trait to students was modeled, described, and then used as a guide for planning all trait-based instruction throughout the year. Each workshop concluded with the collaborative use of planning charts to outline lessons that would be used in upcoming student instruction.

In addition to continuing to address all the writing traits as appropriate, one or two new writing traits were emphasized during each of the three workshops and in the subsequent period of classroom instruction (in addition to the trait of conventions, which was emphasized throughout the year during all workshops, as noted above). During the first one-day workshop, the focus traits were word choice and conventions; the second workshop (held in February) focused on sentence fluency and conventions; the third workshop (held in April) focused on voice, presentation, and conventions.

Of the 103 teachers in the treatment group, 86 teachers (83.5 percent) attended all six training days (the three-day summer institute and all three follow-up workshops). Seventeen teachers (16.5 percent) missed one or two days of training.

Fidelity of implementation

A series of teacher surveys was used to measure the extent to which the instructional practices included in the intervention were applied in the study schools. The items were intended to require only a basic awareness of the general idea of trait-based writing, as would be gained from familiarity with the Oregon statewide standards and assessment system. Teachers completed the survey at three points in time: before the school year in which student data were collected, at midyear, and at the end of the school year. Each survey item required teachers to rate the degree to which they emphasized a particular classroom practice, using a seven-point scale ranging from 0 (“not emphasized at all”) to 6 (“emphasized very often and strongly—very descriptive of my daily classroom”).

Fidelity of implementation in the treatment group was judged at three levels. Teachers who scored above the midpoint (3.0 on a 0–6 scale) on all 10 scales or scored higher than 4.0 on six or more scales were considered to have achieved “advanced” fidelity to the model. Teachers who scored above the midpoint (3.0 on a 0–6 scale) on six or more of the scales but not high enough to achieve advanced implementation were considered to have achieved “basic” fidelity with the model. Teachers who did not reach the basic level of implementation were categorized as “nonimplementers.” These criteria were created for this study based on the recommendations of the developer; their validity has not been established through formal research comparing survey responses to observed measures of classroom practices.

At baseline, 39.6 percent of treatment group teachers and 46.7 percent of control group teachers reported basic or advanced fidelity to the model, with 14.6 percent of treatment group teachers and 16.3 percent of control group teachers reporting advanced fidelity (table 5). These figures suggest that the classroom practices emphasized by the intervention were already in use by many teachers at the outset of the study; this was part of the existing school environment into which the intervention was introduced.

At the time of the midyear survey, 72.1 percent of the treatment group teachers reported a basic or advanced level of fidelity to the model, with 38.1 percent reporting an advanced level of fidelity. At the time of the final survey, 85.6 percent of the treatment group teachers reported a basic or advanced level of fidelity to the model, with 57.8 percent reporting an advanced level of fidelity. At the end of the year, the reported incidence of advanced fidelity among treatment group teachers rose by 43.2 percentage points from levels reported at the beginning of the year; the reported incidence of advanced fidelity among control group teachers rose by 11.2 percentage points.

Table 5. Teacher-reported levels of implementation of classroom practices included in the intervention

	Nonimplementers		Basic fidelity		Advanced fidelity	
Point in time/ group	Number of teachers	Percent of teachers	Number of teachers	Percent of teachers	Number of teachers	Percent of teachers
Baseline[a]						
Treatment	58	60.4	24	25.0	14	14.6
Control	49	53.3	28	30.4	15	16.3
Midyear						
Treatment	27	27.8[b]	33	34.0	37	38.1[b]
Control	47	52.8[b]	27	30.3	15	16.9[b]
End of year[a]						
Treatment	13	14.4[b]	25	27.8	52	57.8[b]
Control	38	41.8[b]	28	30.8	25	27.5[b]

a. A total of four teacher surveys across all three administrations were returned without identifying information and could not be classified by experimental group and were therefore excluded.
b. Differences between the percentage of treatment and control group teachers in these implementation categories were statistically significant (chi-square tests, 1 degree of freedom, $p < .01$).
Source: Authors' analysis, based on data described in text.

At the baseline survey, before any of the study-sponsored professional development had occurred, there were no statistically significant differences between teachers assigned to the treatment condition and teachers assigned to the control group (table 6). Both groups reported similar levels of use of the 10 strategies for writing instruction that would be emphasized in the professional development subsequently provided to the treatment group teachers.

Table 6. Baseline teacher-reported use of instructional strategies for writing

Instructional strategy (number of survey items contributing to the scale)	Coefficient alpha	Mean scale score[a]		Difference	Test statistic[b]
		Treatment group	Control group		
Teaching the language of rubrics for writing assessment (7)	0.91	2.64	2.93	–0.29	$t = -1.31$
		(1.38)	(1.18)		$p = .193$
Reading and scoring papers and justifying the scores (5)	0.79	2.34	2.56	–0.22	$t = -1.19$
		(1.15)	(1.16)		$p = .240$
Teaching focused revision strategies (5)	0.84	3.33	3.61	–0.28	$t = -1.64$
		(1.09)	(1.09)		$p = .105$
Modeling participation in the writing process (4)	0.83	2.54	2.62	–0.08	$t = -0.40$
		(1.25)	(1.37)		$p = .692$
Having students read and analyze materials that demonstrate varying writing quality (3)	0.78	2.39	2.59	–0.20	$t = -0.97$
		(1.24)	(1.23)		$p = .335$
Giving students writing assignments to respond to effective prompts (4)	0.83	3.60	3.60	–0.00	$t = -0.01$
		(1.07)	(1.12)		$p = .993$
Weaving writing lessons into other subjects (4)	0.77	3.11	3.29	–0.18	$t = -1.03$
		(1.13)	(1.12)		$p = .307$
Teaching students to set goals and monitor their progress (5)	0.78	2.85	2.88	–0.03	$t = -0.16$
		(1.16)	(1.19)		$p = .877$
Integrating learning goals for writing into curriculum planning (4)	0.78	3.81	3.76	0.05	$t = 0.33$
		(0.96)	(1.04)		$p = .745$
Teaching ways to structure nonfiction writing (4)	0.80	2.99	3.19	–0.20	$t = -0.98$
		(1.23)	(1.16)		$p = .329$
Total score	0.95	2.96	3.10	–0.14	$t = -0.90$
		(0.96)	(0.96)		$p = .369$

Note: Total score is the mean of the 10 scale scores. Numbers in parentheses are standard deviations.
a. Scores ranged from zero to six.
b. All *t*-tests were calculated accounting for clustering.
Source: Authors' analysis, based on data described in text.

At midyear, in February, teachers responded to the survey again. At this point teachers in the treatment group reported higher levels of use of 9 of the 10 strategies than did control

group teachers (table 7). These differences were statistically significant at $p < .05$; five of the differences were also statistically significant at $p < .01$.

Table 7. Midyear teacher-reported use of instructional strategies for writing

Instructional strategy (number of survey items contributing to the scale)	Coefficient alpha	Mean scale score[a]		Difference	Test statistic[b]
		Treatment group	Control group		
Teaching the language of rubrics for writing assessment (7)	0.92	4.04	2.90	1.14	$t = 5.43$
		(0.98)	(1.25)		$p \leq .001$
Reading and scoring papers and justifying the scores (5)	0.82	3.49	2.58	0.91	$t = 4.26$
		(1.07)	(1.22)		$p \leq .001$
Teaching focused revision strategies (5)	0.88	4.19	3.63	0.56	$t = 2.87$
		(0.96)	(1.11)		$p = .005$
Modeling participation in the writing process (4)	0.85	3.20	2.59	0.61	$t = 2.59$
		1.29	(1.34)		$p = .012$
Having students read and analyze materials that demonstrate varying writing quality (3)	0.67	3.58	2.62	0.96	$t = 5.13$
		(0.96)	(1.17)		$p \leq .001$
Giving students writing assignments to respond to effective prompts (4)	0.80	3.87	3.60	0.27	$t = 1.40$
		(1.03)	(1.09)		$p = .165$
Weaving writing lessons into other subjects (4)	0.76	3.55	3.16	0.39	$t = 2.24$
		(1.06)	(1.08)		$p = .029$
Teaching students to set goals and monitor their progress (5)	0.83	3.34	2.85	0.49	$t = 2.11$
		(1.02)	(1.34)		$p = .038$
Integrating learning goals for writing into curriculum planning (4)	0.81	4.21	3.76	0.45	$t = 2.59$
		(0.91)	(0.99)		$p = .012$
Teaching ways to structure nonfiction writing (4)	0.78	3.47	3.10	0.37	$t = 2.02$
		(1.08)	(1.05)		$p = .047$
Total score	0.95	3.69	3.10	0.59	$t = 3.48$
		(0.83)	(0.97)		$p = .001$

Note: Total score is the mean of the 10 scale scores. Numbers in parentheses are standard deviations.
a. Scores ranged from zero to six.
b. All *t*-tests were calculated accounting for clustering.
Source: Authors' analysis, based on data described in text.

At the end of the school year, after the teachers in the treatment group had been receiving professional development and technical support for about 10 months, teachers in the treatment group reported higher levels of use of all 10 strategies than control group teachers (table 8). These differences were statistically significant at the $p < .01$ level.

Table 8. End-of-year teacher-reported use of instructional strategies for writing

Instructional strategy (number of survey items contributing to the scale)	Coefficient alpha	Mean scale score[a]		Difference	Test statistic[b]
		Treatment group	Control group		
Teaching the language of rubrics for writing assessment (7)	0.93	4.39	3.02	1.37	$t = 6.51$
		(0.92)	(1.28)		$p \leq .001$
Reading and scoring papers and justifying the scores (5)	0.85	3.86	2.74	1.12	$t = 5.00$
		(1.05)	(1.22)		$p \leq .001$
Teaching focused revision strategies (5)	0.90	4.54	3.70	0.84	$t = 4.62$
		(0.91)	(1.13)		$p \leq .001$
Modeling participation in the writing process (4)	0.89	3.55	2.72	0.83	$t = 3.72$
		(1.25)	(1.35)		$p \leq .001$
Having students read and analyze materials that demonstrate varying writing quality (3)	0.80	3.86	2.81	1.05	$t = 5.17$
		(1.03)	(1.14)		$p \leq .001$
Giving students writing assignments to respond to effective prompts (4)	0.82	4.32	3.77	0.55	$t = 2.84$
		(0.94)	(1.13)		$p = .006$
Weaving writing lessons into other subjects (4)	0.81	4.05	3.36	0.69	$t = 3.81$
		(0.95)	(1.17)		$p \leq .001$
Teaching students to set goals and monitor their progress (5)	0.85	3.74	2.99	0.75	$t = 3.15$
		(1.10)	(1.29)		$p = .002$
Integrating learning goals for writing into curriculum planning (4)	0.84	4.40	3.89	0.51	$t = 2.99$
		(0.86)	(1.06)		$p = .004$
Teaching ways to structure nonfiction writing (4)	0.82	3.93	3.37	0.56	$t = 2.90$
		(1.06)	(1.11)		$p = .005$
Total score	0.96	4.06	3.24	0.82	$t = 4.60$
		(0.84)	(1.02)		$p \leq .001$

Note: Total score is the mean of the 10 scale scores. Numbers in parentheses are standard deviations.
a. Scores ranged from zero to six.
b. All *t*-tests were calculated accounting for clustering.
Source: Authors' analysis, based on data described in text.

Teachers in both groups reported similar levels of use of these classroom strategies at baseline; at the end of the year, treatment group teachers reported greater use of all 10 strategies than did the control group teachers. However, the control group did report some use of these strategies, and the magnitude of the difference between the treatment group teachers and the control group teachers at the end of the year is difficult to interpret. From baseline to the end of the year, the treatment group teachers' average total score increased 1.10 points on the scale, while the control group teachers' average total score increased 0.14 points. At the end of the year, the treatment group teachers' average total score of 4.06 corresponded to the response scale anchor statement that the classroom strategies were "emphasized often or strongly," while the control group teachers' average total score of 3.24 corresponded to the portion of the scale between the responses of "emphasized often or strongly" and "emphasized somewhat." The extent to which this difference in reported practices reflects a thorough implementation by treatment group teachers or a lack of implementation by control group teachers cannot be determined from the data collected during the study. These data provide information about how one group compares to the other group, but do not provide definitive information about the absolute level of implementation by either group.

Notwithstanding the difficulties in interpreting the data from this self-report survey, it appears that the study did not capture a contrast between schools that fully implemented the treatment versus schools that did not implement it at all. This is expected in a large-scale effectiveness study in which implementation is not tightly controlled. At baseline, 46.7 percent of control group teachers reported basic or advanced fidelity to the model; this proportion rose to 47.2 percent at midyear and 58.3 percent at the end of the year, suggesting that the practices promoted by the 6+1 Trait Writing model were used in some control group classrooms during the year. The corresponding proportion of treatment group teachers—those who reported basic or advanced fidelity to the model—rose from 39.6 percent at baseline to 72.1 percent at midyear and 85.6 percent at the end of the year. Conversely, at midyear 27.8 percent of treatment group teachers reported being nonimplementers, and at the end of the year 14.4 percent of treatment group teachers still reported being nonimplementers. The estimated impact on students, reported in the next chapter, must be interpreted within this context.

These findings are based on the self-reported practices of teachers in the treatment and control group schools. No additional measurements using more complex methods—such as multiple independent classroom observations—were conducted to corroborate or validate these teacher reports. It is possible that teachers may have over- or under-reported the extent to which they implemented the practices that were the focus of the surveys, and it is possible that teachers in one experimental group may have systematically under- or over-reported more or less than teachers in the other experimental group. In order to address this problem and minimize the tendency for participants to distort their responses in ways they considered to be socially desirable (to please the researchers or to appear to be "good" teachers), teacher reports were collected using paper surveys rather than interviewer-administered surveys, and teachers were assured that the research methods would protect their privacy (for example, school administrators did not have access to individual survey answers). Still, the extent to

which teachers may have biased their responses is unknown, since other measures of their classroom practices were not obtained.

Although most of the questions did not ask specifically about the use of writing traits, 13 of the 45 survey questions referred to the use of writing traits in general, without referencing details of the 6+1 Trait Writing intervention. Because writing traits are part of the Oregon standards and assessments, it was assumed that teachers in both the control and treatment groups would understand the general concept of writing traits and would be able to answer questions about whether they used them in their instruction. Nevertheless, to determine whether the exclusion of items that referred to writing traits might affect the implementation fidelity findings, the survey scores were recalculated without these items. Details of this analysis are presented in appendix I.

Excluding these results had no effect on the results at baseline except that the scale for "teaching the language of rubrics" no longer existed because all items on that scale referred to writing traits. At midyear, differences between control and treatment groups on two scales ("teaching focused revision strategies" and "modeling participation in the writing process") were no longer significant after dropping items that referred to writing traits. Teachers in the treatment group still reported higher levels of use of six of the nine remaining strategies and had significantly higher total survey scores than the control group teachers. At the end of the year, the exclusion of these items had no effect on the survey results except that the scale for "teaching the language of rubrics" no longer existed. Differences between treatment and control group teachers remained significant on all nine remaining scales.

4. Results of analysis for the confirmatory question

This chapter presents the results of analyses performed in order to answer the confirmatory question: What is the impact of 6 + 1 Trait Writing on grade 5 student achievement in writing? The chapter begins with the results of the impact analysis used to estimate the treatment effect. This is followed by the results of sensitivity analyses conducted to determine whether alternative ways of modeling the data would change the substantive results of the impact analysis. The complete multilevel model results for the confirmatory question are in appendix J.

Impact analysis

The outcome measure for the confirmatory research question was a score for overall writing quality, on a six-point scale ranging from 1 (indicating a severely flawed essay demonstrating very little or no mastery) to 6 (indicating an essay demonstrating clear and consistent mastery of writing). Table 9 shows the unadjusted estimated means for the pretest and posttest scores, disaggregated by the treatment conditions. The standard errors of the estimates were calculated to reflect the nesting of students within schools.

Table 9. Unadjusted estimated pretest and posttest means and standard errors for the treatment and control groups

	Pretest		Posttest	
Group	Mean	Standard error	Mean	Standard error
Treatment	3.600	0.032	3.927	0.044
Control	3.664	0.034	3.905	0.047

Note: Sample size (including imputed values for missing posttest scores) was 2,230 for the treatment group and 1,931 for the control group.
Source: Authors' analysis, based on data described in text.

The difference in the mean posttest score between the treatment and the control groups in table 9 does not provide the most accurate available estimate of the treatment effect. A more accurate estimate of the treatment effect involves taking into consideration pretest scores and the matched pairing of schools before random assignment, which was based on school rates of eligibility for free or reduced-price lunch. These covariates were included in the full statistical model, along with additional covariates that adjusted for exogenous baseline differences between schools on three teacher-reported measures: the school average for the weekly teacher-reported hours students spend in class practicing writing, the school average for teacher years of teaching experience, and the school average for teacher years of experience teaching writing.

The estimated treatment effect is illustrated in table 10 as the difference between the covariate-adjusted posttest scores of treatment group and control group students. The standard error of the estimate is adjusted for the nesting of students within schools, as are the significance test and the confidence interval for the difference. The estimated

treatment effect was 0.100 on the original scale, which has a range of 1–6 in half-point increments. The difference was statistically significant and resulted in a standardized treatment effect of 0.109. Approximately 5.5 percent of the student records were missing posttest scores; in accordance with the original analysis plan, missing values were estimated using multiple imputation before this benchmark analysis was conducted. Two of the baseline teacher-reported covariates were found to be significantly associated with the outcome among paired schools. Years of teaching experience was negatively associated with the outcome ($z = -2.38$, $p = .017$). Years of teaching writing was positively associated with the outcome ($z = 2.00$, $p = .045$).

Table 10. Estimated impact of 6+1 Trait Writing adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

Outcome measure: adjusted posttest score	Treatment group (n = 2,230)	Control group (n = 1,904)	Difference: treatment effect (SE)	Summary treatment effect[c] (SE)	Test statistic	95% confidence interval	Effect size[d]
Paired[a]	3.743	3.613	0.130 (0.046)	0.100 (0.044)	$z = 2.27$ $p = .023$	0.014 to 0.185	0.109
Singleton[b]	3.848	4.131	−0.283 (0.162)				

a. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in the referent pair, under the condition that the school aggregates of the exogenous teacher variables were set to zero.
b. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in which no students were eligible for free or reduced-price lunch, under the condition that the school aggregates of the exogenous teacher variables were set to zero.
c. The summary treatment effect is a pooled estimate combining the treatment effect among paired schools (those randomly assigned within matched pairs based on district and percentage of students eligible for free or reduced-price lunch) and singleton schools (those randomly assigned to experimental condition after all other schools in their district had been assigned as part of matched pairs).
d. Glass's *delta* (standardized difference using the control group standard deviation of the posttest scores). See appendix K.
Source: Authors' analysis, based on data described in text.

Sensitivity analysis

Two sensitivity analyses were conducted. The first of these involved replicating the impact analysis after deleting data from students who did not complete a posttest, rather than imputing their posttest scores. (With the exception of this sensitivity test, all other analyses in this report were performed on the full dataset including imputed values.) The second sensitivity analysis replicated the benchmark statistical model except for the exclusion of the three covariates that had been used to adjust for baseline differences in school averages for teacher-reported hours students spend in class practicing writing, teacher years of teaching experience, and teacher years of experience teaching writing.

Replicating the impact analysis after dropping students without posttest scores. For this test, the impact analysis was replicated after removing from the sample those students who did not complete the posttest. (Aside from this sensitivity test, all other analyses in this report were performed on the full dataset including imputed values). The results, reported in table 11, were similar to those of the benchmark analysis. The standardized effect size was 0.110, compared to the effect size of 0.109 in the benchmark analysis.

Table 11. Estimated impact of 6+1 Trait Writing adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect: Excluding students without posttest scores

Outcome measure: adjusted posttest score	Treatment group (n = 2,114)	Control group (n = 1,817)	Difference: treatment effect (SE)	Summary treatment effect[c] (SE)	Test statistic	95% confidence interval	Effect size[d]
Paired[a]	3.737	3.611	0.126 (0.044)	0.100 (0.042)	z = 2.37 p = .018	0.017 to 0.183	0.110
Singleton[b]	3.865	4.139	−0.274 (0.166)				

a. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in the referent pair, under the condition that the school aggregates of the exogenous teacher variables were set to zero.
b. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in which no students were eligible for free or reduced-price lunch, under the condition that the school aggregates of the exogenous teacher variables were set to zero.
c. The summary treatment effect is a pooled estimate combining the treatment effect among paired schools (those randomly assigned within matched pairs based on district and percentage of students eligible for free or reduced-price lunch) and singleton schools (those randomly assigned to experimental condition after all other schools in their district had been assigned as part of matched pairs).
d. Glass's *delta* (standardized difference using the control group standard deviation of the posttest scores). See appendix K.
Source: Authors' analysis, based on data described in text.

Model without school-level covariates to control for exogenous differences between treatment and control teachers at baseline. Another sensitivity analysis was performed to determine whether using a model that did not statistically control for exogenous differences found between treatment and control teachers at baseline would change the result of the impact analysis. The following covariates from the baseline teacher survey were removed from the benchmark model: the school average for teacher years of teaching experience, the school average for teacher years of experience teaching writing, and the school average for the weekly teacher-reported hours students spend in class practicing writing. The substantive result remained the same as that of the impact analysis, in that a positive significant impact of the treatment was found (table 12). The standardized effect size was 0.081, compared to the effect size of 0.109 in the main analysis.

Table 12. Estimated impact of 6+1 Trait Writing adjusted for pretest score and pair fixed effect

Outcome measure: adjusted posttest score	Treatment group (n = 2,230)	Control group (n = 1,931)	Difference: treatment effect (SE)	Summary treatment effect[c] (SE)	Test statistic	95% confidence interval	Effect size[d]
Paired[a]	3.735	3.638	0.097 (0.039)	0.074 (0.037)	z = 1.98 p = .048	0.001 to 0.148	0.081
Singleton[b]	3.980	4.106	−0.126 (0.118)				

a. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in the referent pair.
b. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in which no students were eligible for free or reduced-price lunch.
c. The summary treatment effect is a pooled estimate combining the treatment effect among paired schools (those randomly assigned within matched pairs based on district and percentage of students eligible for free or reduced-price lunch) and singleton schools (those randomly assigned to experimental condition after all other schools in their district had been assigned as part of matched pairs).
d. Glass's *delta* (standardized difference using the control group standard deviation of the posttest scores). See appendix K.
Source: Authors' analysis, based on data described in text.

5. Results of additional exploratory analyses

This chapter presents the results of exploratory analyses of whether 6+1 Trait Writing had an impact on grade 5 student achievement in particular writing traits and whether the intervention had differential effects based on student gender or ethnicity. The complete multilevel model results for exploratory analyses are in appendix L.

Results on measures of particular writing traits

Table 13 compares the covariate-adjusted posttest scores on each of the trait scales for the treatment and control groups. The difference between the two groups represents the treatment effect. The standard error of the estimate is adjusted to reflect the nesting of students within schools.

For all six trait scales, the estimated mean for the treatment group exceeded that for the control group. (The trait of presentation was not scored or analyzed, because students did not have the opportunity to address presentation issues during the assessment.) Three of these differences reached statistical significance at $p < .05$. For these three writing traits (organization, voice, and word choice) effect sizes ranged from 0.117 to 0.144. For the other three traits (ideas, sentence fluency, and conventions) the mean outcome score of students in the treatment condition was higher than that of students in the control condition, but these differences were too small to be considered statistically significant.

Table 13. Estimated impact of 6+1 Trait Writing on individual trait scale scores, adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

Outcome measure: adjusted posttest score	Treatment group (n = 2,230)	Control group (n = 1,931)	Difference: treatment effect (SE)	Summary treatment effect[c] (SE)	Test statistic	95% confidence interval	Effect size[d]
Ideas							
Paired[a]	4.176	4.070	0.106 (0.041)	0.039 (0.038)	z = 1.03 p = .302	−0.035 to 0.113	0.070
Singleton[b]	4.462	4.802	−0.340 (0.097)				
Organization							
Paired[a]	3.953	3.873	0.080 (0.029)	0.060 (0.028)	z = 2.15 p = .031	0.005 to 0.114	0.117
Singleton[b]	4.055	4.335	−0.280 (0.116)				
Voice							
Paired[a]	4.403	4.311	0.092 (0.028)	0.062 (0.027)	z = 2.28 p = .023	0.009 to 0.116	0.132
Singleton[b]	4.429	4.761	−0.332 (0.103)				
Word Choice							
Paired[a]	4.069	3.983	0.086 (0.025)	0.055 (0.023)	z = 2.37 p = .018	0.009 to 0.101	0.144
Singleton[b]	4.128	4.298	−0.170 (0.068)				

Sentence Fluency							
Paired[a]	4.008	3.917	0.091 (0.032)	0.053 (0.030)	$z = 1.77$ $p = .076$	−0.006 to 0.111	0.112
Singleton[b]	4.138	4.385	−0.247 (0.088)				
Conventions							
Paired[a]	3.980	3.914	0.066 (0.035)	0.028 (0.033)	$z = 0.864$ $p = .388$	−0.036 to 0.092	0.054
Singleton[b]	4.062	4.276	−0.214 (0.090)				

Note: All the trait scale score outcome measures are covariate adjusted.
a. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in the referent pair, under the condition that the school aggregates of the exogenous teacher variables were set to zero.
b. The covariate-adjusted posttest score for the treatment (control) group represents the predicted posttest score of a student who had a pretest score at the grand mean and who attended a treatment (control) school in which no students were eligible for free or reduced-price lunch, under the condition that the school aggregates of the exogenous teacher variables were set to zero.
c. The summary treatment effect is a pooled estimate combining the treatment effect among paired schools (those randomly assigned within matched pairs based on district and percentage of students eligible for free or reduced-price lunch) and singleton schools (those randomly assigned to experimental condition after all other schools in their district had been assigned as part of matched pairs).
d. Glass's *delta* (standardized difference using the control group standard deviation of the posttest scores). See appendix K.
Source: Authors' analysis, based on data described in text.

Did the results vary by student gender or ethnicity?

The second exploratory question asked whether student gender or ethnicity affected the impact of 6+1 Trait Writing on grade 5 student achievement. This section presents the results, which show that neither factor had a statistically significant effect. As with the confirmatory analysis, students who were English language learners and who were being pulled out of the regular classroom for specialized writing instruction were not included in the sample. Therefore, the findings cannot be applied to this group. English language learners whose English language proficiency allowed them to participate in the mainstream classroom during writing instruction were included in the sample.

In the tables and discussion below, the difference in the treatment effect between two subgroups is referred to as a "moderator effect"—meaning that subgroup membership could potentially "moderate" or "modify" the treatment effect—and this effect is

represented in the statistical analysis by the interaction between a student's treatment status and his or her subgroup status.

Girls versus boys. The estimated treatment effect for boys was larger than that for girls, although the difference was not statistically significant (table 14). The subgroup-by-treatment cell means are presented in table 15.

Table 14. Estimated moderator effect of gender on holistic scores, adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

Outcome measure: adjusted posttest score	Treatment effect for girls (n = 2,112)	Treatment effect for boys (n = 2,044)	Difference: moderator effect (SE)	Summary moderator effect (SE)	Test statistic	95% confidence interval
Paired	0.084	0.167	0.083 (0.058)	0.096 (0.053)	z = 1.83 p = .067	−0.007 to 0.199
Singleton	−0.352	−0.201	0.151 (0.120)			

Note: Gender codes were originally missing for five students, but these gender codes were filled in using imputed values, so the analysis was based on a sample size of 4,161.
Source: Authors' analysis, based on data described in text.

Table 15. Estimated holistic score means for girls and boys in the treatment and control groups, adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

	Treatment group		Control group	
Gender	Mean	Standard error	Mean	Standard error
Girls (n = 2,112)	3.835	0.112	3.828	0.113
Boys (n = 2,044)	3.684	0.112	3.579	0.111

Note: Each mean score is covariate adjusted. Gender codes were originally missing for five students, but these gender codes were filled in using imputed values, so the analysis was based on a sample size of 4,161.
Source: Authors' analysis, based on data described in text.

White non-Hispanics versus all others. When the ethnic categories were defined as White non-Hispanics versus all others, there was no significant difference in the treatment effect between the two groups (table 16).

Table 16. Estimated moderator effect of White non-Hispanic/other ethnicity on holistic scores, adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

Outcome measure: adjusted posttest score	Treatment effect for White non-Hispanics (n = 3,163)	Treatment effect for all others (n = 998)	Difference: moderator effect (SE)	Summary moderator effect (SE)	Test statistic	95% confidence interval
Paired	0.121	0.152	0.031 (0.069)	0.002 (0.066)	z = 0.03 p = .980	−0.128 to 0.131
Singleton	−0.247	−0.549	−0.302 (0.222)			

Note: Ethnicity codes were originally missing for 10 students. The ethnicity of these students was coded as "unknown," and they were included as a part of "all others." Consequently, the analysis was based on a sample size of 4,161.
Source: Authors' analysis, based on data described in text.

The subgroup-by-treatment cell means are presented in table 17.

Table 17. Estimated holistic score means for White non-Hispanics and all others in treatment and control groups, adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

	Treatment group		Control group	
Ethnic group	Mean	Standard error	Mean	Standard error
White non-Hispanics (n = 3,163)	3.772	0.114	3.713	0.114
All others (n = 998)	3.725	0.120	3.670	0.120

Note: Each mean score is covariate adjusted. Ethnicity codes were originally missing for 10 students. The ethnicity of these students was coded as "unknown," and they were included as a part of "all others." Consequently, the analysis was based on a sample size of 4,161.
Source: Authors' analysis, based on data described in text.

Taken together, the results of this set of exploratory subgroup analyses suggest that the impact of 6+1 Trait Writing on student achievement did not have differential effects on White and non-White students.

White non-Hispanics versus Hispanics. When the ethnic categories were defined as White non-Hispanics versus Hispanics, there was no significant difference in the treatment effect between the two groups (table 18). Students of other ethnic backgrounds were not included in this analysis.

Table 18. Estimated moderator effect of White non-Hispanic/Hispanic ethnicity on holistic scores, adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

Outcome measure: adjusted posttest score	Treatment effect for White non-Hispanics (n = 3,163)	Treatment effect for Hispanics (n = 516)	Difference: moderator effect (SE)	Summary moderator effect (SE)	Test statistic	95% confidence interval
Paired	0.120	0.177	0.057 (0.093)	0.026 (0.089)	z = 0.29 p = .769	−0.148 to 0.201
Singleton	−0.252	−0.542	−0.290 (0.300)			

Source: Authors' analysis, based on data described in text.

The subgroup-by-treatment cell means are presented in table 19.

Table 19. Estimated holistic score means for Hispanics and White non-Hispanics in treatment and control groups, adjusted for exogenous differences in schools at baseline, pretest score, and pair fixed effect

	Treatment group		Control group	
Ethnic group	Mean	Standard error	Mean	Standard error
Hispanic (n = 516)	3.732	0.127	3.648	0.132
White non-Hispanic (n = 3,163)	3.804	0.116	3.742	0.116

Note: Each mean score is covariate adjusted.
Source: Authors' analysis, based on data described in text.

Taken together, the results of this set of exploratory subgroup analyses suggest that the 6+1 Trait Writing treatment did not have a differential effect on student achievement among White non-Hispanic students and Hispanic students.

6. Summary of findings and study limitations

This section summarizes the study findings regarding the effect of 6+1 Trait Writing on grade 5 student achievement in writing. It also identifies the study's limitations.

The study was designed to estimate the impact of 6+1 Trait Writing on student achievement in writing during the first year of a typical implementation. This question was addressed among grade 5 students in Oregon using a single "holistic" writing score on student essays. Exploratory analyses using six scores on specific traits of writing were also conducted. Professional development was provided by the model developers in the same year that student assessments were administered. The particulars of school and classroom implementation of the approach were allowed to vary in the schools, without any special oversight or intervention by the developers beyond the technical assistance normally offered to those who receive the materials and professional development.

Effect of 6+1 trait writing on student achievement

For the confirmatory research question (What is the impact of 6+1 Trait Writing on grade 5 student achievement in writing), use of the 6+1 Trait Writing model caused a statistically significant difference in student writing scores, with an effect size of 0.109 (p = .023). This means that the estimated average score of students in the treatment group was 0.11 standard deviations higher than the estimated average score of students in the control group.

Another way of understanding effect size is in terms of improvement in percentile scores. An intervention with an effect size of 0.11 would increase the average level of achievement from the 50th to the 54th percentile. (For a table of effect sizes and corresponding increases from percentile scores of 50, see www.bestevidence.org/methods/effectsize.htm.)

The effect size that is derived from a particular experimental study is dependent on several factors, including the reliability and precision of the outcome measure used, any additional factors that are observed and accounted for in statistical models, and the "cause size" or strength of the treatment that is tested.

Approximately 5.5 percent of the student records were missing posttest scores. For this primary or "benchmark" analysis of the data, in accordance with the original plan for the study, missing values were estimated using a statistical procedure called multiple imputation before the analysis was conducted. This allowed the preservation of the full randomized sample of students, rather than removing a non-random subset of students (those with missing data points) from the sample.

Additional analyses were conducted in order to determine whether the finding from the benchmark analysis was sensitive to variations in the method used for the statistical modeling of the data. In both of these sensitivity tests the substantive result of the study remained consistent with the benchmark analysis.

The first sensitivity analysis was conducted after those students with missing data were removed from the sample. For the benchmark analysis, posttest scores had been imputed based on other data for those students who did not complete the posttest. This method has the advantage of preserving the original randomized sample. However, another possible way to analyze the data is to restrict the sample to those students who completed both pretests and posttests. This approach can introduce selection bias due to differences in attrition in the treatment and control groups. The analyses of this subset of students within the data resulted in an effect size of 0.110 (p = .018). Aside from this sensitivity test, all other analyses in this report were performed on the full dataset including imputed values.

Another sensitivity analysis was conducted using only the student pretest score and the indicator of how schools were matched as covariates, without adjusting for exogenous baseline differences in schools as reported by teachers. Teachers in treatment and control schools reported differences at baseline on several characteristics. Control teachers reported more years of teaching experience and more years of experience teaching writing when compared to treatment teachers. Control teachers also reported that their students spent fewer hours per week practicing writing in class. The model had been statistically adjusted for these differences in the benchmark analysis but was not adjusted in the second sensitivity analysis. The analysis without adjustment for baseline difference in these teacher-reported measures resulted in an estimated effect size for the treatment of 0.081 (p = .048).

Two exploratory research questions were also addressed. The first examined the impact of 6+1 Trait Writing on student achievement in particular traits of writing. The primary analysis model was repeated six additional times to estimate the impact of the intervention on each of the six traits of writing. Use of the 6+1 Trait Writing model caused a statistically significant difference in three writing traits — organization, voice, and word choice — with effect sizes ranging from 0.117 to 0.144 (p = .031 to .018). Although the mean outcome score of students in the treatment condition was higher than that of students in the control condition for the other three traits — ideas, sentence fluency, and conventions — the differences were too small to be considered statistically significant given the size and sensitivity of the experiment.

The second exploratory research question examined whether there may be differences in impact according to the gender or ethnicity of students. No differential effects of the intervention were found based on student ethnicity or gender.

Limitations of the study

Progress in science involves an accumulation of knowledge from many similar, repeated studies conducted under slightly different circumstances. A single study rarely provides a definitive or conclusive answer to a question of interest. This experimental study contributes to the research on writing instruction, but the findings reported here are limited by a number of contextual factors:

- The intervention studied in this research was a first-year implementation of the 6+1 Trait Writing model that provided additional writing instruction and assessment strategies intended to complement whatever writing curricula and instructional strategies were present in the participating schools. Questions about the interaction of the model with any specific writing curriculum were not addressed and cannot be answered using these findings. Likewise, questions about curriculum materials designed to fully integrate a trait-based approach to writing were not addressed by this research. The findings reported here cannot be applied to answer such questions.
- The implementation of recommended classroom strategies by the treatment group and control group teachers was measured using newly developed self-report surveys that have not been validated by observational or other measures. These surveys provided information about implementation by one group relative to the other group and were subject to possible biases in teacher self-reports of their classroom practices. The extent to which the model was actually implemented by treatment group teachers is unknown, as is the extent to which treatment group teachers implemented these strategies more than they were implemented by the control group teachers.
- The findings reported are for grade 5 students in 74 Oregon schools that volunteered to participate. The extent to which these findings apply to other grade levels, other schools, or other regions is unknown.
- The extent to which the findings would be replicated in other settings, such as targeted implementations for particular schools or student populations, is unknown and cannot be inferred from these results.
- The student achievement data were collected during the same school year in which teachers received their first year of professional development in the 6+1 Trait Writing model. The study does not answer questions about what effects might be produced by longer durations of professional development and/or classroom implementation.
- It is possible that teachers or students in the treatment group may have responded differently to the knowledge that they were participating in an experimental study than did teachers or students in the control group; if so, any difference or lack of difference in the performance of teachers or students in the two groups could have been due in part to this differential response to participation in a research study.

Appendix A. Professional development institute and workshop objectives and agendas

The goal of the 6+1 Trait Writing model is to integrate effective writing assessment with classroom instruction to improve student writing. The model provides teachers and students with a common framework and language that is based on the seven "traits" of effective writing: ideas, organization, voice, word choice, sentence fluency, conventions, and presentation (see box 1 in the main report). Classroom teachers are provided with a set of specific classroom practices that help them select learning activities to engage students in learning about and using the traits to self-evaluate their writing. These practices include teaching students the language of rubrics for the traits, having students score and justify their scores on writing samples, and using a writing process that emphasizes feedback, revision, and editing. The approach is focused on using formative writing assessments to provide effective and timely feedback on student writing performance.

Professional development and resources followed the format indicated by the developers, who recommend a three-day summer institute followed by three full-day workshops strategically scheduled throughout the school year. Teachers received a participant manual, information updates through a monthly newsletter, and online technical assistance from the 6+1 Trait Writing professional development team.

Summer institute

The three-day summer institute provided teachers with an overview of the writing process, a description of the analytic assessment method for providing effective teacher feedback to students, and strategies for teaching students to self-evaluate their writing. The agenda for the summer institute is outlined in table A1. The institute provided opportunities for group discussion and hands-on activities in which teachers practiced using an analytic rubric or scoring guide and shared ideas for integrating writing in their classroom instruction. The primary objectives of the summer institute were to:

- Provide teachers with an understanding of the traits and increase their skills in using the analytic rubric to provide feedback to students about their writing.
- Familiarize teachers with the components of an effective student writing process.
- Acquaint teachers with the 10 classroom strategies for successful integration of the traits in the classroom.

Teachers received information about classroom strategies and technical assistance to plan lessons and student activities for integrating traits—particularly ideas and organization—in their classrooms from September through November.

Table A1. Agenda for the three-day 6+1 Trait Writing Summer Institute

Day	Agenda
1	1. General introduction and warm-up activity to connect teachers to the day's agenda. 2. Review of the history of the 6+1 Trait Writing model. 3. Overview of the analytic scoring rubrics. a. Group practice using the analytic scoring rubrics to score student papers. b. Review of components of the student writing process. 4. Introduction of 10 classroom strategies used to teach students the traits one by one.
2	1. Continued discussion of 10 classroom strategies to teach students the traits one by one. 2. Modeling and sharing of trait-based activities and sample lessons using the 10 classroom strategies. 3. Group and individual practice scoring student papers using the analytic scoring rubrics. 4. Review of the seven traits and planning.
3	1. Continued discussion of use of the 10 strategies to teach students the traits one by one. 2. Modeling and sharing of trait-based activities and lessons using the 10 classroom strategies. 3. Group and individual practice scoring student papers using the analytic scoring rubrics 4. Discussion of how to create effective writing prompts. 5. Review of all traits and plan lessons to teach ideas and organization traits from September through November.

Source: 6+1 Trait Writing Summer Institute training agendas and records.

Participant manual. Every teacher attending the summer institute received a participant manual, which provided information about the 6+1Trait Writing model, an overview of the analytic writing assessment tools, and descriptions of the seven traits. The manual also included sample lesson plans, student activity ideas, and classroom teaching resources for implementing the 10 classroom strategies. The manual provided teachers with lists of books, teaching suggestions, and sample forms to help them implement trait-based writing instruction in their classrooms.

Learning the traits. Throughout the three-day institute, teachers were provided with information and engaged in hands-on activities to increase their understanding of the seven writing traits. Teachers developed an understanding of each trait by reading a definition of the trait and participating in activities involving the trait, such as reading writing examples aloud or participating in a simulated trait lesson. Teachers then practiced scoring and providing feedback on sample student papers. These practice opportunities were followed by whole group discussion of the scoring activity to increase understanding of the assessment process and scoring consistency among the training participants.

Ten classroom strategies. Teachers were introduced to the 10 classroom strategies that were designed to teach students about each of the traits, provide analytic feedback to students about their writing, and engage students in editing their own writing (box A1). Examples of these strategies include teaching the language of the traits, having students practice using the analytic rubrics to score their papers, structuring peer editing

opportunities, and integrating lessons on each trait across different content areas. During each day of the institute, the trainer reviewed the traits, engaged teachers in analytic scoring practices, and modeled lesson plans that teachers could use in their classrooms.

For example, the trainer made picture books available for teachers to review and explained how the books could be used to provide students with examples of particular writing traits. The trainer modeled how to use these books to introduce and teach students about the traits. The trainer also showed teachers how to create effective writing prompts for grade 5 students. Teachers were provided opportunities to practice generating writing prompts to use in their classrooms and to plan how they would integrate trait-based instruction in their daily classroom instruction during the first three months of school.

Box A1. Ten trait-based classroom strategies

1. Teach students the language they need to speak and think like writers. This practice can be accomplished in a number of ways: rewriting the rubric, teaching the vocabulary, using the rubrics to explain each indicator.

2. Read, score, and justify (RSJ) your scores on anonymous sample papers. This practice can be accomplished by having students score model papers as a group and then asking them to justify the score given by using the rubric of choice. (This activity also reinforces teaching the language of the rubric.)

3. **Practice focused revision steps** by selecting a weak paper and then revising it as a group by

- Working with a partner or small group.
- Working on an anonymous weak paper selected by the teacher.
- Revising for one trait or descriptor at a time. This strategy also reinforces and teaches the writing process, reinforcing the connection between the writing process and the traits.

4. Model writing right now! This means you! Write along with your students. The model doesn't have to be a student's assignment; let them see you as a writer. Take a risk and share your "work in progress" with students. Ask students for revising feedback. You'll be amazed! This practice consists of a teacher simply writing a piece alongside the students and then allowing students to score the teacher's work, give feedback, and perhaps even revise the piece as a group.

5. Read, read, read, and read some more. Read printed material of all kinds to illustrate strengths and weaknesses in writing. This practice consists of exposing students not just to the five modes of writing (descriptive, narrative, expository, persuasive, and imaginative) but also of providing them with a variety of materials to read, such as menus, video game directions, driver's manuals, grocery lists, directions for building cabinets, greeting cards, and so forth.

6. Practice structures of writing/ modes. The organizational arrangement of content within a written piece varies, based on its purpose, and should be closely tied to the prompts. This practice provides graphic organizers to students for use with each mode, thereby providing a skeleton structure of written organization by mode. Students use the graphic organizers to construct their written pieces in ways that enrich, clarify, and organize their writing to better communicate the content by mode.

7. Use effective writing prompts. The MC*RAFTS (Mode, Context * Role, Audience, Format, Topic, Strong verb) strategy builds on a teaching strategy known as the RAFT technique (Santa et al. 1988), adding an M for mode, a C for context, and an S for strong verb. The practice provides a prompt to students, which gives them a guide for their written responses. Students apply higher levels of Bloom's taxonomy as they provide written responses that address the RAFTS. Using this technique, teachers create thoughtful, explicit writing prompts that encourage students to write in a variety of modes for different audiences. The "S" (strong verb) component helps students determine the mode of writing to be used. Students use classroom strategy 6 and the graphic organizer that goes with it. MC*RAFT prompts also connect what students know and are learning in other content areas (writing across content areas) to various modes of writing and the related structures of writing (classroom strategy 6)

8. Set goals and monitor progress. Teach students to set writing goals and continuously help them monitor their progress. Using the analytic rubrics that accompany this model of writing instruction, students learn how to self-assess their progress and monitor their growth. This practice results in timely, meaningful feedback and student self-assessment—best practices that result in improved student achievement

9. Organize activities and mini-lessons by trait. Weave focused trait skill lessons into your curriculum to enhance your writing program. (This section of the manual provides an index of the lessons, trait by trait.) This classroom practice helps teachers organize the resources they have used in their classrooms to teach writing, as well as providing the nearly 400-page "text" book for guidance throughout the year.

10. Plan and implement the traits. Use planning charts to plan and decide where the traits best fit in your writing program, throughout the year. This classroom practice uses the rubric to plan the writing instruction. Teachers break down the rubric line by line, turning each row into a student objective. Once a trait unit has been properly taught, students will have received instruction on everything expected of them according to the rubric used. Once students understand what is expected of them and can self-assess and monitor their progress, their writing improves.

Source: 6+1 Trait Writing Summer Institute training materials.

Planning to incorporate ideas and organization. During the third day of the summer institute, teachers used the planning chart to systematically integrate trait-based instruction in their classrooms. A primary objective was for each teacher to leave with formal lesson plans to integrate ideas and organization into their classroom instruction and student activities before the first follow-up meeting, in November.

For each planning session, teachers used the "trait cycle of instruction" to organize lesson plans and classroom activities that focus on the traits students were to learn and use according to the professional development timeline. The steps of this planning process include the following:

1. Use the 6+1 Trait rubric to write lesson plans for each trait.
2. Teach the trait language by rewriting the rubric in student-friendly language.
3. Practice scoring sample student papers, justify the scores by using the appropriate rubric, and discuss the justification for each score.
4. Develop trait-based activities and lesson plans to teach the particular trait. (During the workshop, the trainer modeled a variety of trait-based activities and lesson plans that teachers could use in their planning process
5. Practice creating effective writing prompts. Gather written products to conduct formative assessment on student writing. Measure the overall growth of student as well as student improvement in each specific trait.

Follow-up workshops. Teachers attended three full-day follow-up workshops, in November, February, and April. The objective of these meetings was to allow teachers time away from their classrooms to discuss issues related to implementing trait-based writing, to increase their understanding of how to implement the 6+1 Trait model, and to help them plan lessons for integrating specific traits into their classroom instruction. The follow-up workshops also provided teachers with the opportunity to discuss their progress toward implementing the model by identifying what was working and what was not and by generating ideas to address specific problems related to model implementation or student writing. The timeline for the introduction of each trait is outlined in table A2.

Table A2. Timeline for planning classroom implementation of writing traits

Trait	Three-day summer institute	November	February	April
Ideas	**X**			
Organization	**X**			
Word choice		**X**		
Conventions	**X**	**X**	**X**	**X**
Sentence fluency			**X**	
Voice				**X**
Presentation				**X**

Source: 6+1 Trait Writing Summer Institute training agendas and records.

Each workshop followed the same general agenda outlined in box A2. The agenda included a troubleshooting session during which teachers discussed their challenges and successes in using the 6+1 Trait model during the previous months and shared ideas about how to improve writing instruction in their classrooms. The trainer provided a quick review of all of the traits and the analytic assessment process, with special emphasis on the traits teachers were expected to implement in the following months. Teachers were provided opportunities to practice scoring sample student papers and to discuss the rationale for their scoring decisions.

Box A2. Agenda for the one-day 6+1 Trait Writing workshops

1. General introduction and warm-up activity to connect teachers to the day's agenda.

2. Group discussion of what's working, what's not working. Use problem solving to list possible interventions for each identified problem.

3. Review the 6+1 Trait® Writing model:

- Seven traits.
- Components of the student writing process.
- 10 classroom strategies.
- Trait cycle of instruction.

4. Review the analytic scoring rubric, and practice scoring sample student papers.

5. Model how to use the trait cycle of instruction to plan activities, practice scoring, plan lessons, and write effective prompts.

6. Plan lesson plans and student activities to teach and provide opportunities to practice trait-based instruction.

7. Closure and evaluation.

Source: 6+1 Trait Writing Summer Institute training agendas and records.

During each follow-up workshop, two or three writing traits were the focus of the planning session. To help teachers plan lessons and activities, the trainer described and modeled how to use the cycle of instruction for introducing the new traits to students. Each workshop concluded with time for teachers to plan lessons and activities for the upcoming school months.

Online support resources. Teachers were provided online support resources and information about the 6+1 Trait model. The online system provided teachers with sample grade 5 student papers that illustrated specific traits, additional opportunities to practice scoring using the analytic rubric and compare their scores to those of expert raters, and suggested books and reading materials to illustrate particular traits of writing. Teachers also had access to technical assistance from the project trainers through email correspondence and telephone conversations.

Monthly newsletter. Teachers received a monthly newsletter that provided information about the traits they were expected to incorporate in their classrooms as well as information regarding writing instruction resources. The newsletter also provided teachers with information about upcoming professional development activities to encourage attendance and engagement in the writing model.

Appendix B. Statistical power analysis

This appendix describes the statistical power analysis that was used during the planning phase of the study in order to determine the sample size and the expected minimum detectable effect size. To determine the level of statistical power attainable from various sample sizes, the authors performed a set of analyses to estimate the minimum detectable effect size under several scenarios. The minimum detectable effect size was defined as the necessary size of effect in order to maintain statistical power of 0.8. The power analysis was performed to detect the main effect of the intervention on the student outcome.

In collaboration with the Oregon Department of Education, the authors calculated the intraclass correlation coefficient (ICC) for a two-level cluster randomized trial (CRT) model very similar to that proposed for the current study (student nested within school). The ICC was calculated using the 2005/06 grade 4 Oregon Writing Assessment data (n = 39,057). First, the unconditional ICC was calculated by fitting a two-level CRT model without any covariate. The conditional ICC was then calculated by fitting a two-level CRT model with the previous year's building mean as the school-level (L2, for level 2) covariate. The effect of this covariate, R^2_{L2}, was calculated from the school-level variance in the two models. Specifically:

- When a two-level CRT model without the covariate was fit to the data, school-level variance (τ) was 2.917 and student-level variance (σ^2) 18.221. The unconditional ICC was therefore calculated as $\tau/(\tau + \sigma^2) = 0.138$.
- When a two-level CRT model with the covariate was fit to the data, school-level conditional variance ($\tau_{|x}$) was 1.574 and student-level variance (σ^2) was 18.225. The conditional ICC was therefore calculated as $\tau_{|x}/(\tau_{|x} + \sigma^2) = 0.079$.

The effect of the covariate, R^2_{L2}, was calculated from the unconditional school-level variance (τ) and the school-level variance conditional on the use of a covariate ($\tau_{|x}$):

$R^2_{L2} = 1 - (\tau_{|x}/\tau) = 0.428$.

The previous year's school mean (school-level covariance) and the true school mean were correlated, $R_{L2} = 0.654$.

Based on this information, a power analysis was performed using the Optimal Design software, with an unconditional ICC of 0.14 and an effect of the covariate (R^2_{L2}) of 0.43. The number of schools required to attain the minimum detectable effect (MDE) size of 0.10, 0.15, 0.20, and 0.25 are reported in table B1.

Table B1. Estimated minimum detectable effect size given varying sample sizes

Number of schools	Minimum detectable effect size (at power of 0.8)
320	0.10
144	0.15
84	0.20
54	0.25

Note: Unconditional ICC is 0.14, R^2_{L2} is 0.43.
Source: Authors' analysis, based on data described in text.

The general goal of the power analysis was to estimate the number of schools needed for the sample in order to maintain power of 0.8 for a minimum detectable effect size of $\delta = 0.25$. The number of students per teacher (n) was set at 20. The number of teachers per school (J) was set at 2. The default value of 0.05 was used for the alpha level. This analysis was an approximation, as it did not take into account the individual-level baseline writing measure as a covariate, but rather took into account its school-level aggregate. The degree to which an individual-level covariate increases power at the school level was not modeled in the Optimal Design software available at the time.

The power analysis indicated that a minimum of 54 schools would be needed in order to ensure adequate statistical power to detect a difference of 0.25 standard deviations or greater between the treatment and control schools. Initial recruitment plans called for a sample of 64 schools, each with a minimum of 30 students in grade 5, in order to accommodate possible school or student attrition. However, during early discussions with Oregon school districts it became clear that some districts would participate only if all of their schools, including those with fewer than 30 grade 5 students, could be involved in the study. To accommodate the district requests to include small schools in the study while still preserving the desired level of statistical power, the final sample size included 74 schools, each with a minimum of 20 students in grade 5.

The power analysis was recalculated in July 2008 to determine whether the estimate of minimum detectable effect size changed based on the possibility that some schools might include fewer than 20 grade 5 students after attrition and to examine the updated sample size of 74 schools. All the input variables described above remained the same, except that the number of schools was increased to 74 and the number of students per school was varied across three levels representing the smallest schools. The calculation was performed three times, with the number of students per school set to 15, 20, and 30. The resulting MDEs were $\delta = 0.25$ with 15 students per school, $\delta = 0.23$ with 20 students per school, and $\delta = 0.22$ with 30 students per school. The increased school-level sample size appeared to be successful in offsetting the loss of power caused by the fact that some schools had small numbers of grade 5 students.

A retrospective power analysis was conducted in March 2011 in order to determine the achieved statistical sensitivity of the design based on the input values calculated from the actual data. Table B2 displays these figures for the main analysis as well as the exploratory analyses of individual scale scores. The retrospective power analysis was based on the original planned benchmark model, which did not control for exogenous teacher variables. This allows for "apples-to-apples" comparisons of predicted versus actual statistical sensitivity.

Table B2. Actual minimum detectable effect sizes

	Analysis model for pairs		Analysis model for singletons		Pooled analysis
Analysis	Conditional ICC	MDES	Conditional ICC	MDES	MDES
Research question 1, impact on holistic score	0.015	0.13	0.043	0.40	0.16
Exploratory analyses of specific writing traits					
Ideas	0.022	0.14	0.055	0.44	0.17
Organization	0.006	0.11	0.060	0.45	0.13
Voice	0.016	0.14	0.086	0.52	0.16
Word choice	0.019	0.14	0.039	0.39	0.17
Sentence fluency	0.015	0.14	0.043	0.40	0.17
Conventions	0.020	0.14	0.035	0.37	0.17

Source: Authors' analysis, based on data described in text.

Appendix C. Test administration directions and sample student essay booklet

This appendix contains the test administration directions and a sample student essay booklet.

Test administration directions

General guidelines

- The writing assessments will be completed over three consecutive days during 45- to 50-minute uninterrupted periods. Please notify Education Northwest if there are problems that may interfere with this assessment schedule.
- Teachers will give brief instructions to students each assessment day, but students will be asked to write on their own, without peer or teacher assistance in planning, drafting, revising, or editing.
- Feel free to help individual students by rereading the prompt to them, but do not offer suggestions about what to write and do not proofread or otherwise edit the student work.
- Most students will use some of the first session and part of the second to prewrite/brainstorm and write rough drafts. The remainder of the second session is often used to revise and edit rough drafts. The third session is often used to complete the revision and copy final papers.
- As much as possible, each student should be allowed to proceed at his or her own pace. Students who finish ahead of others should have reading materials or other planned activities available so that they will not disturb those who need additional time.
- All study group students need to complete the writing assessment in English.

Resources

- Students may use a dictionary and thesaurus of the type normally available in your school as resources for word definitions, usage, or spelling.
- Students may NOT use handouts or locally developed handouts that go beyond word definitions, usage, or spelling guides; they may not use reference sources (textbooks, encyclopedias, or almanacs) or peer editing.
- Students should have access or use accommodations or modifications only if it is part of their IEP or 504 Plan.
- Students may not use a computer or word processor unless it is specified in their IEP or 504 Plan.

Administering the student writing assessment

1. Recheck the Writing Booklets and read the directions

- Before class, check to make sure each grade 5 student in the study group has a Student Writing Project booklet (Writing Booklet) with his or her name on the cover page.
- If you have a fifth grade student who is in the study group and was not assigned a Writing Booklet, write his or her name on the Classroom Roster and the cover sheet of the Writing Booklet that has the SAME code number that is next to the student's name on the Classroom Roster.
- Please review the "General Writing Assessment Guidelines" and the following writing assessment instructions before beginning the assessment.

2. Administer the Student Writing Assessment – DAY 1

- Hand out the Writing Booklets. Make sure all students clear off their desks and have a pen or pencil.
- Read the information in the boxes verbatim to all students.

This week we are going to work on a writing assessment that will be sent to Portland for scoring. Please do not write your name on the booklet – your name is already written on the cover page. For this writing assessment, you will be asked to write on your own, without help from me (teacher) or other students. Also, you will not be allowed to take your writing assessment home. You will be given three days to complete the writing assessment.

STEP ONE: Planning
Look in your writing booklet. You must write on the topic printed on pages 2 and 3 of your booklet. You may use the planning page on page 3 of your booklet to list ideas, or do some other prewriting BEFORE you write your rough draft.

STEP TWO: Writing the rough draft
Begin writing your rough draft on pages 4 and 6 of your booklet when you finish your prewriting. Please note that the draft and final pages are next to each other to make it easier for you to write your final copy.

STEP THREE: Revising, editing, and writing your final copy
When your rough draft is finished, you should spend some time revising and editing. You may use any of the editing tools we use in the classroom to edit your paper, except help from me (teacher) or other students. When you are done revising and editing, recopy your paper onto the Final Copy pages in your booklet (pages 5 and 7). Please make your final copy as neat as you can so that it is easy for others to read.

Please do not take your booklets home—they should be turned into me at the end of the writing period.

3. Collect the booklets

- Collect the booklets at the end of the assessment period and store in a secure place.

Administering the student writing assessment – days 2 and 3

4. Continue the writing assessment on days 2 and 3.

- Hand out the student booklets. Make sure all students clear off their desks and have a pen or pencil. Read the information in the box verbatim to all students.

> Today, you are going to (continue/complete) work on your writing assessment. Remember, you may use any of the editing tools we use in the classroom to edit your paper, except help from me (teacher) or other students. Please use the planning, rough draft, and final copy pages in your booklet. As a reminder:
>
> **STEP ONE:** Planning
> Look in your writing booklet. You must write on the topic printed on pages 2 and 3 of your booklet. You may use the planning page on page 3 of your booklet to list ideas, or do some other prewriting BEFORE you write your rough draft.
>
> **STEP TWO:** Writing the rough draft
> Begin writing your rough draft on pages 4 and 6 of your booklet when you finish your prewriting. Please note that the draft and final pages are next to each other to make it easier for you to write your final copy.
>
> **STEP THREE:** Revising, editing, and writing your final copy
> When your rough draft is finished, you should spend some time revising and editing. You may use any of the editing tools we use in the classroom to edit your paper, except help from me (teacher) or other students. When you are done revising and editing, recopy your paper onto the Final Copy pages in your booklet (pages 5 and 7). Please make your final copy as neat as you can so that it is easy for other to read.
>
> Please do not take your booklets home – turn them into me at the end of the period.

5. Day 2 – Collect the booklets

- Collect the booklets at the end of the assessment period, and store in a secure place.

6. Day 3 – Return the booklets to the Site Coordinator.

- When the writing assessment is completed, check all booklets to see that the student completed the assessment and that the drafts and final copies are on the designated pages. If not, add explanatory notes on the final copy page.
- In the Spring Posttest Column on the Classroom Roster, complete the Status, Accommodations, and Modifications columns as follows:

  Status: Enter the pretest status for each student using the Status Codes at the bottom of the Classroom Roster page. Enter code "1-Assessment completed" for students who completed the assessment. For students who did not complete the assessment, enter the code that best describes the reason for the incomplete.

Accommodations and Modifications: Write **Y** if the student received accommodations or modifications during the Fall pretest and **N** if the student did not receive accommodations or modifications.

- Check to ensure that each student has an entry for grade, sex, and ethnicity on the Classroom Roster.
- Put the Classroom Roster and all Writing Booklets (complete and incomplete) assigned to study group students in the Classroom Packet. Return the Classroom Packet to your school Site Coordinator.

Sample student essay booklet

Name: _____

Grade: _____

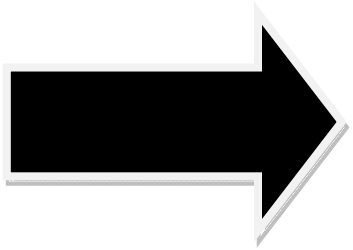
Teacher: _____

Student Writing Project

Code: 281521059

Code: 281521059

Continue to next page to start your writing project.



INSTRUCTIONS

1. In this writing assignment you are asked to describe and explain something. Below is your topic; please read it carefully:

   Think of a skill you have learned that has made your life easier or more fun. Write a letter telling about your skill, **explain** how you learned it, and why you think it is important.

2. Use the next page, called **PLANNING**, to plan and organize your ideas.

3. Write your draft on the pages called **DRAFT**.

4. Mark your revisions and editing changes on the same pages called **DRAFT**.

5. Write your finished copy in BLACK INK on the pages called **FINAL COPY**.

Think of a skill you have learned that has made your life easier or more fun. Write a letter telling about your skill, **explain** how you learned it, and why you think it is important.

PLANNING

DRAFT

Page 1

Write your draft on lines below.

FINAL COPY

Page 1

Please write in BLACK PEN

DRAFT

Page 2

Write your draft on lines below.

FINAL COPY

Page 2

Please write in BLACK PEN

Continue FINAL COPY on this page if you need more space.

Do not write on this page.

Appendix D. Teacher survey

The following pages reproduce the survey that was administered to teachers three times during the study. Table D1 provides information about the numbered survey items that were included on each of the scales that are presented in the "Fidelity of implementation" section of the main report.

Table D1. Teacher survey items included in scale scores

Instructional strategy scale	Survey items included on scale
Teaching the language of rubrics for writing assessment	Items 6, 7, 19, 26, 29, 37, 45
Reading and scoring papers and justifying the scores	Items 1, 10, 22, 30, 43
Teaching focused revision strategies	Items 8, 21, 25, 33, 38
Modeling participation in the writing process	Items 2, 11, 20, 34
Having students read and analyze materials that demonstrate varying writing quality	Items 9, 18, 31
Giving students writing assignments to respond to effective prompts	Items 17, 28, 35, 41
Weaving writing lessons into other subjects	Items 5, 12, 13, 42
Teaching students to set goals and monitor their progress	Items 16, 36, 39, 40, 44
Integrating learning goals for writing into curriculum planning	Items 3, 14, 23, 32
Teaching ways to structure nonfiction writing	Items 4, 15, 24, 27

Source: Authors' records.

Teacher Survey on Writing Instruction

The Northwest Regional Educational Laboratory (NWREL) developed this survey to better understand how teachers currently teach writing and what classroom activities students engage in that pertain to writing. There are no right answers—except honest and accurate ones! The purpose is NOT to evaluate you as a teacher but to better understand what teachers really do in their classrooms.

Confidentiality: Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

SECTION 1: YOUR WRITING INSTRUCTION last semester, during Winter 2009

For the following questions, please rate your level of emphasis on each of these potential instructional strategies in your classroom. For each question, circle the number that best describes your classroom practices **last semester, during Winter 2009**, using the rating scale below:

0	1	2	3	4	5	6
Not emphasized at all		Emphasized somewhat		Emphasized often or strongly		Emphasized very often and strongly–very descriptive of my daily classroom

1	In my classroom, students evaluated writing passages from literature as part of learning how to think about and discuss writing.	0	1	2	3	4	5	6
2	I used examples of my own writing when teaching students about writing.	0	1	2	3	4	5	6
3	I planned my class so that students had time and support for the writing process (e.g., drafting, revising, publishing).	0	1	2	3	4	5	6
4	I provided my students examples of effective non-fiction writing using different structures (e.g., sequential, cause and effect, problem and solution)	0	1	2	3	4	5	6
5	In my classroom, students received detailed feedback and scores on their writing as part of assignments in other subject areas (e.g., science, math, social studies).	0	1	2	3	4	5	6
6	I used trait language in lessons.	0	1	2	3	4	5	6
7	In my classroom, students used "trait vocabulary" appropriately across the curriculum.	0	1	2	3	4	5	6
8	In my classroom, students used concepts and language about the traits of writing while revising their writing or responding to the writing of others.	0	1	2	3	4	5	6

9	In my classroom, students actively engaged in critiquing the materials we read in class using trait criteria.	0	1	2	3	4	5	6
10	In my classroom, students spent time discussing and justifying the scores given to particular writing passages.	0	1	2	3	4	5	6
11	I reflected aloud on my own writing, using trait vocabulary to show students how to think about writing.	0	1	2	3	4	5	6
12	I explained the specific writing criteria that are important for each subject area (e.g., science, math, social studies).	0	1	2	3	4	5	6
13	I integrated significant writing tasks with student assignments in other subject areas (e.g., science, math, social studies).	0	1	2	3	4	5	6
14	In my instructional planning, I targeted specific learning outcomes for students in the standard mechanics of English writing.	0	1	2	3	4	5	6
15	In my classroom, students practiced writing lead sentences and paragraphs, body paragraphs, and conclusions for a variety of non-fiction purposes.	0	1	2	3	4	5	6
16	In my classroom, students participated in real publishing opportunities (e.g., writing competitions, commercial publications, school-wide newsletters).	0	1	2	3	4	5	6
17	When I gave writing assignments to students, I used focused prompts that clearly identified the audience and purpose for the writing.	0	1	2	3	4	5	6
18	I gave my students reading assignments that taught them to identify how effective writing differs from ineffective writing.	0	1	2	3	4	5	6
19	In my classroom, students talked about the traits of writing.	0	1	2	3	4	5	6
20	I modeled for students how to receive feedback and reflect on my own writing, using trait criteria.	0	1	2	3	4	5	6
21	I used trait concepts and language when providing feedback to students to help them revise their writing.	0	1	2	3	4	5	6
22	In my classroom, students used analytic scoring guides to evaluate their own papers.	0	1	2	3	4	5	6
23	I targeted specific learning outcomes for aspects of writing other than mechanics.	0	1	2	3	4	5	6
24	In my classroom, students practiced constructing thesis statements for non-fiction writing (e.g. expository and persuasive).	0	1	2	3	4	5	6
25	As part of my writing instruction, I taught specific strategies to revise initial drafts into more polished final versions.	0	1	2	3	4	5	6
26	Trait definitions and age-appropriate rubrics were readily available and/or posted in my classroom.	0	1	2	3	4	5	6
27	In my classroom, students practiced non-fiction writing using a variety of structures (e.g., sequential, cause and effect, problem and solution)	0	1	2	3	4	5	6
28	I gave students writing assignments that required them to write for a variety of purposes (e.g., expository, persuasive, narrative).	0	1	2	3	4	5	6
29	I used trait language in giving students feedback about their writing.	0	1	2	3	4	5	6
30	In my classroom, students used an analytic scoring guide to evaluate a variety of writing forms (e.g., posters, leaflets, letters, essays).	0	1	2	3	4	5	6

31	I read examples of effective writing within various subject areas (e.g., science, math, social studies) and discussed writing skills.	0	1	2	3	4	5	6
32	I continuously assessed students' writing and provided feedback on their progress in learning writing skills.	0	1	2	3	4	5	6
33	In my classroom, students spent time revising their writing using trait criteria, as a separate, conscious step in the writing process.	0	1	2	3	4	5	6
34	I invited student comment on my writing	0	1	2	3	4	5	6
35	I gave my students opportunities to write in a variety of forms (e.g., essays, posters, presentations, brochures).	0	1	2	3	4	5	6
36	In my classroom, students kept track of how their individual writing skills developed over time.	0	1	2	3	4	5	6
37	In my classroom, students talked about their own writing or that of others using trait concepts and language.	0	1	2	3	4	5	6
38	In my classroom, students used a writing process (e.g., draft, revise, publish).	0	1	2	3	4	5	6
39	In my classroom, students used scores on their writing to identify goals to improve their writing.	0	1	2	3	4	5	6
40	I provided a systematic way for students to store and organize their writing.	0	1	2	3	4	5	6
41	In my classroom, students were asked to write for a wide variety of audiences (e.g., other students, newspaper readers, other cultures).	0	1	2	3	4	5	6
42	I used mini-lessons to review important writing skills	0	1	2	3	4	5	6
43	Examples of student writing were displayed around my classroom and used as part of classroom instruction.	0	1	2	3	4	5	6
44	I encouraged students to actively seek feedback on their writing.	0	1	2	3	4	5	6
45	I communicated the trait model of writing to parents and community members.	0	1	2	3	4	5	6

46. What writing program do you currently use in your teaching? _____

_____

_____

47. How many hours per week on average do your students practice their writing in class? _____

48. How many hours per week on average do your students spend on homework that includes significant writing?

_____

49. How do you grade your students' writing assignments? Please circle the number before one answer:

1. Single grade for whole assignment
2. Single grade for whole assignment with feedback comments
3. Separate grades for different writing skills
4. Separate grades for different writing skills with feedback comments

SECTION 2: YOUR PROFESSIONAL BACKGROUND

50. What is your highest degree? Please circle the number before one answer:

1. B.A. or B.S.
2. M.A. or M.S.
3. Ph.D. or Ed.D.
4. Other (Please describe) _____

51. Which of the following most accurately describes the type of teaching credential you currently hold?

Please circle the number before one answer:

1. A **regular** or **standard** state certificate
2. An **emergency certificate** or **waiver** that is issued for a specified time period
3. Other (Please describe) _____

52. Counting this year, how many years have you been a full-time classroom teacher?

53. Counting this year, how many years have you been teaching writing? _____

54. Please list any training you have received in the last two years related to teaching writing:

_____

_____

55. How well prepared do you believe you are to teach writing?

1	2	3	4	5
Not at all prepared	Only a little prepared	Fairly well prepared	Very well prepared	Extremely well prepared

56. How confident are you to teach writing?

1	2	3	4	5
Not at all confident	Only a little confident	Fairly confident	Very confident	Extremely confident

Your Name: _____

School: _____

Thank you for completing this survey.

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless such collection displays a valid OMB control number. The valid OMB control number for this information collection is 1850-0835. The time required to complete this form is estimated to average 30 minutes per respondent, including the time to review instructions and complete the survey. If you have any comments concerning the accuracy of the time estimate or suggestions for improving this form, please write to: U.S. Department of Education, Washington, DC 20202-4651. If you have any comments or concerns regarding the status of your individual submission of this form, write directly to: Dr. Michael Coe, Northwest Regional Educational Laboratory, 101 SW Main, Suite 500, Portland, OR 97204.

Appendix E. Scoring rubrics for student essays

This appendix describes the two types of study rubrics applied by the scoring teams: the holistic rubric and the six analytic rubrics.

Holistic rubric

Score of 6. An essay in this category demonstrates *clear and consistent mastery*, although it may have a few minor errors. A typical essay

- Effectively and insightfully addresses the topic and task and demonstrates outstanding development of a theme, using clearly appropriate supporting elements to elaborate the theme.
- Is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas.
- Exhibits skillful use of language, using a varied, accurate, and apt vocabulary
- Demonstrates meaningful variety in sentence structure.
- Is free of most errors in grammar, usage, and mechanics.

Score of 5. An essay in this category demonstrates *reasonably consistent mastery*, although it will have occasional errors or lapses in quality. A typical essay

- Effectively addresses the topic and task and demonstrates strong development of a theme, generally using appropriate supporting elements to elaborate the theme.
- Is well organized and focused, demonstrating coherence and progression of ideas.
- Exhibits facility in the use of language, using appropriate vocabulary.
- Demonstrates variety in sentence structure.
- Is generally free of most errors in grammar, usage, and mechanics.

Score of 4. An essay in this category demonstrates *adequate mastery*, although it will have lapses in quality. A typical essay

- Addresses the topic and task and demonstrates competent development of a theme, using adequate supporting elements to elaborate the theme.
- Is generally organized and focused, demonstrating some coherence and progression of ideas.

- Exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary.
- Demonstrates some variety in sentence structure.
- Has some errors in grammar, usage, and mechanics.

Score of 3. An essay in this category demonstrates *developing mastery* and is marked by ONE OR MORE of the following weaknesses:

- Addresses the topic and task and demonstrates some development of a theme but may do so inconsistently or using inadequate supporting elements to elaborate the theme.
- Is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas.
- Displays developing facility in the use of language but sometimes uses weak vocabulary or inappropriate word choice.
- Lacks variety or demonstrates problems in sentence structure.
- Contains an accumulation of errors in grammar, usage, and mechanics.

Score of 2. An essay in this category demonstrates *little mastery* and is flawed by ONE OR MORE of the following weaknesses:

- Addresses the topic and task in a way that is vague or seriously limited, and demonstrates weak development of a theme, using inadequate supporting elements to elaborate the theme.
- Is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas.
- Displays very little facility in the use of language, using very limited vocabulary or incorrect word choice.
- Demonstrates frequent problems in sentence structure.
- Contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured.

Score of 1. An essay in this category demonstrates *very little* or *no mastery*, and is severely flawed by ONE OR MORE of the following weaknesses:

- Barely addresses the topic and task and demonstrates little or no development of a theme, using very inadequate supporting elements to elaborate the theme.
- Is disorganized or unfocused, resulting in a disjointed or incoherent essay.

- Displays fundamental errors in vocabulary.
- Demonstrates severe flaws in sentence structure.
    - Contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.

Essays not written on the essay assignment will receive a score of zero.

Analytic rubrics six-trait/six-point analytical scoring guide

Ideas/content (development)

6 **This paper is extremely clear or focused. Relevant anecdotes and details enrich the central theme.** A. The topic is narrow and manageable. B. Relevant, telling, quality details give the reader important information that goes beyond the obvious or predictable. C. Accurate, precise details are present to support the main ideas; appropriate use of resources provides strong, accurate, credible support. D. The writer seems to be writing from knowledge or experience; the ideas are fresh and original. E. The reader's questions are anticipated and answered. F. The writing makes connections and shares insights, an understanding of life, and a knack for picking out what is significant.	**5** **The ideas/content in this piece are well marked by detail and information.** A. The topic is focused but still could use additional narrowing. B. More than half the time the details and support are clear and relevant. Other details are general but stay with the topic. C. Credible details are present which support the main idea/theme. D. Some new ways of thinking about this topic are presented. E. The writer is clearly aware of questions the reader may have and attempts to answer them. F. A clear theme has been developed from the topic.	**4** **The writer has defined the topic, although the development is basic or general**. A. The topic is fairly broad; however, it is clear where the writer is headed. B. Support is attempted, but doesn't go far enough yet in fleshing out the key issues or story line. C. Ideas are reasonably clear, though they may not be detailed, personalized, accurate, or expanded enough to show in-depth understanding or a strong sense of purpose. D. A few examples of "showing" are present, but the writer relies on general examples. E. The reader is left with a few questions but is generally clear about the content. F. The writer stays on the topic and begins to develop a theme.
3 **The reader can understand the main ideas, although they may be broad or simplistic.** A. The topic is becoming clear; however, because it is so broad or lacks specific focus, the reader often must infer to get the overall message. B. Support is sporadic. C. A general sense of the idea is present though not enhanced by significant details. D. A heavy reliance on "telling," not "showing" examples E. The reader is left with many questions due to lack of specific information. F. The writer has not yet focused the topic past the obvious.	**2** **No one main idea stands out yet, although possibilities are emerging.** A. The paper hints at topics, but doesn't settle on one yet. B. Support is incidental or confusing. C. Several possible ideas may be present which could become central themes/ideas on different pieces of writing. D. The writer makes statements without specifics to back them up. E. The reader has so many questions because of the lack of specific information. It is hard to "fill in the blanks." F. Glimmers of the writer's topic or main point show up occasionally.	**1** **As yet, the paper has no clear sense of purpose or central theme. To extract meaning from the text, the reader must make inferences based on sketchy or missing details. The writing reflects more than one of these problems:** A. The writer is still in search of a topic, brainstorming, or has not yet decided on the main idea of the piece. B. Information is limited or unclear or the length is not adequate for development. C. The idea is a simple restatement of the topic or an answer to the question with little or no attention to detail. D. The writer has not begun to define the topic in a meaningful, personal way. E. Everything seems as important as everything else; the reader has a hard time sifting out what is important. F. The text may be repetitious or may read like a collection of disconnected, random thoughts with no discernable point.

Organization

6 **The organization enhances and showcases the central idea or theme. The order, structure, or presentation of information is compelling and moves the reader through the text**.

A. An inviting introduction draws the reader in; a satisfying conclusion leaves the reader with a sense of closure and resolution.
B. Thoughtful transitions clearly show how ideas connect.
C. Details seem to fit where they're placed; sequencing is logical and effective.
D. Pacing is well controlled; the writer knows when to slow down and elaborate, and when to pick up the pace and move on.
E. The title, if desired, is original and captures the central theme of the piece.
F. Organization flows so smoothly the reader hardly thinks about it; the choice of structure matches the purpose and audience.

5 **The organization is smooth with only a few small bumps here and there**.

A. The writer goes farther than the obvious beginning and conclusion, but needs to step up one more notch.
B. The transitions are logical but may lack originality.
C. Sequencing makes sense and moves a step beyond the most obvious structure.
D. Though the pacing is under control, there are still places the writer needs to highlight or move through more quickly.
E. The title (if required) settles for a key idea rather than capturing a deeper theme.
F. The organization generally works satisfactorily if not yet so smooth to escape obvious detection.

4 **The organizational structure is strong enough to move the reader through the text without too much confusion**.

A. The paper has a recognizable introduction and conclusion. The introduction may not create a strong sense of anticipation; the conclusion may not tie up all loose ends.
B. Transitions often work well; at other times, connections between ideas are fuzzy.
C. Sequencing shows some logic, but not under control enough that it consistently supports the ideas. In fact, sometimes it is so predictable and rehearsed that the structure takes attention away from the content.
D. Pacing is fairly well controlled, though the writer sometimes lunges ahead too quickly or spends too much time on details that do not matter.
E. A title (if desired) is present, although it may be uninspired or an obvious restatement of the prompt or topic.
F. The organization sometimes supports the main point or storyline; at other times, the reader feels an urge to slip in a transition or move things around.

3 **The organization is somewhat problematic and slows the readers ability to engage in the text**.

A. Either the intro or conclusion or both are clichés or just leave you wanting a lot more.
B. Transitions, when present, are repetitive or misleading.
C. The structure has taken over so completely, it dominates the ideas. The sequencing is painfully obvious.
D. The writer lets one part of the piece dominate and loses control over the pacing,
E. There is just a passing glimmer of how the title (if desired) was selected for this piece.
F. The organization of the piece begins to distract from the content.

2 **The organization of the piece needs a great deal of work to be effective. Only moments here and there give the writer a clue about what's going on**.

A. The lead and/or conclusions are ineffective to guide the readers.
B. A little bit of help is offered to get from one idea to the next but not often enough to keep the reader from being confused.
C. So little useful structure is present, it's hard to get a picture of how the piece fits together as a whole.
D. Pacing feels awkward; the writer slows to a crawl when the reader wants to get on with it, and vice versa.
E. A title (if desired) doesn't match the content.
F. The organization is often problematic and frustrates the reader as they struggle with the ideas.

1 **The writing lacks a clear sense of direction. Ideas, details, or events seem strung together in a loose or random fashion; there is no identifiable internal structure. The writing reflects more than one of these problems:**

A. There is no real lead to set up what follows, no real conclusion to wrap things up.
B. Connections between ideas are confusing or not even present.
C Sequencing needs lots and lots of work to make sense.
D. Pacing is not yet being considered.
E. No title is present (if requested.)
F. Problems with organization make it hard (almost impossible) for the reader to get a grip on the main point or story line.

Voice

6 The writer speaks directly to the reader in a way that is individual, compelling and engaging. The writer "aches with caring," yet is aware and respectful of the audience and the purpose for writing. A. The reader feels a strong interaction with the writer, sensing the person behind the words. B. The writer takes a risk by revealing who they are and what they think. C. The tone and voice give flavor and texture to the message and are appropriate for the purpose and audience. D. Narrative writing seems honest, personal, and written from the heart. Expository or persuasive writing reflects a strong commitment to the topic by showing why the reader needs to know this and why they should care. E. This piece screams to be read aloud, shared, and talked about. The writing makes you think about and react to the author's point of view. F. The writing shows control and consistency in its use of voice throughout.	**5 A sincere attempt has been made to address the purpose and audience for the writing in an interesting way. It skips a beat here and there, however.** A. It's a strong attempt although the best moments fade in and out. B. Moments of insight make this piece come alive. C. The writer pays attention to which tone is best used on this piece. It's not totally consistent but leans in the right direction. D. Narrative writing has many moments when the writer feels connected. E. Expository or persuasive writing leaves the reader with a sense of why the writer chose these ideas. F. The voice is strong throughout the pieces, but the writer slacks off a bit here and there.	**4 The writer seems sincere, but not fully engaged or involved. The result is pleasant or even personable, but not compelling**. A. The writing communicates in an earnest, pleasing manner. B. Only one or two moments here or there surprise, delight, or move the reader. C. The writer seems aware of an audience but weighs ideas carefully and discards personal insights in favor of safe generalities. D. Narrative writing seems sincere, but not passionate; expository or persuasive writing lacks consistent engagement with the topic to build credibility. E. The writer's willingness to share his/her point of view may emerge strongly at some places, but is often obscured behind vague generalities. F. The reader senses the voice the writer was striving for, but must rely on their own intuition to "read it in" rather than the writer being in control of the voice.
3 It would be hard to point to a unique moment or two, although the reader is trying desperately to "hear" the writer. A. The writer keeps the reader a safe distance away. Hope of connecting is all that keeps the reader going. B. No special moments stand out. It's all pretty much the same. C. It's more important for this writer to hide and be safe than to try and connect. D. Narrative writing tells only what it must. No care is shown to help the writer feel anything. E. The reader has to wonder if the writer cares one way or the other about that topic. (Expository of persuasive.) F. A glimmer of voice is all that is found here and that's a generous reading.	**2 The voice in the piece relies on the readers good faith to hear or feel anything in phrases such as "I like it" or "It was fun."** A. The writing sits on the surface and doesn't reach out past the most clichéd of phrases. Yawn. B. The writing is humdrum and "risk-free." C. The writer doesn't acknowledge the needs of the reader to understand any point of view in the piece. D. Narrative writing is just an outline and doesn't have any detail to engage the reader. E. As an expository or persuasive piece it lacks any conviction or authority to distinguish it from a mere list of facts. F. So many chances and yet the writer misses every opportunity to engage the reader.	**1 The writer seems indifferent, uninvolved, or distanced from the topic and/or the audience. As a result, the paper reflects more than one of the following problems:** A. The writer speaks in a kind of monotone that flattens all potential highs or lows of the message. B. The lack of voice begins to lull the reader to sleep. C. The writer is not concerned with the audience, or the writer's style is a complete mismatch for the intended reader. D. The writing is lifeless or mechanical; depending on the topic, it may be overly technical or jargonistic. E. Narrative? Expository? Who can tell? F. No point of view is reflected in the writing—zip, zero, zilch, nada.

Word choice

6 Words convey the intended message in a precise, interesting, and natural way. The words are powerful and engaging. A. Words are specific and accurate; it is easy to understand just what the writer means. B. The words and phrases create pictures and linger in your mind. C. The language is natural and never overdone; both words and phrases are individual and effective. D. Striking words and phrases often catch the reader's eye—and linger in the reader's mind. (You can recall a handful as you reflect on the paper.) E. Lively verbs energize the writing. Precise nouns and modifiers add depth and specificity. F. Precision is obvious. The writer has taken care to put just the right word or phrase in just the right spot.	**5 Attempts are made to reach for better and more precise words although not as often as possible.** A. Words are correct and in many cases they are "just right." B. It's easy to understand what they writer is communicating. Several "mind pictures" are present. C. As the writer tries new words and phrases, they are usually more right than wrong. D. The verbs are more active but still may need a little attention here and there. E. There's care and attention paid to selecting the best words to fit the piece. It's moved past the "just functional stage." F. The words and phrases are working really well.	**4 The language is functional, even if it lacks much energy. It is easy to figure out the writer's meaning on a general level.** A. Words are adequate and correct in a general sense; they simply lack much flair and originality. B. Familiar words and phrases communicate, but rarely capture the reader's imagination. Still, the paper may have one or two fine moments. C. Attempts at colorful language show a willingness to stretch and grow, but sometimes it goes too far (thesaurus overload!). D. The writing is marked by passive verbs, everyday nouns and adjectives, and lack of interesting adverbs. E. The words are only occasionally refined; it's more often, "the first thing that popped into my mind." F. The words and phrases are functional—with only a moment or two of sparkle.
3 The language is interpretable but without any energy. A little interpretation is needed to understand some parts. A. Words are mostly adequate but add no flavor to the piece. B. Simple words are all that are attempted and they may be so general they distract from the meaning. The verbs lack any pizzazz. C. Few attempts are made at colorful or figurative language and even those work only at a limited level. D. Although most of the parts of speech can be identified in the sentence, some misuse is confusing to the reader. E. The words feel like a rote response and reflect a lack of craftsmanship. F. The reader gets meaning from the words in only the most general way.	**2 So many places are flawed that meaning is often impaired. Wrong words are used and the reader can't see any connection to the idea being shared.** A. Language is so vague (e.g., It was a fun time, She was neat, It was nice, We did lots of stuff) that only a limited message comes through. B. Even simple words are used incorrectly. The verbs if present are flat. C. No attempts are made to use figurative or colorful language. D. Limited vocabulary and/or frequent misuse of parts of speech impair understanding. E. Jargon or clichés distract or mislead. Persistent redundancy distracts the reader. F. If you work very hard you can get a general understanding of what the piece is about - but it's not easy.	**1 The writer struggles with a limited vocabulary, searching for words to convey meaning. The writing reflects more than one of these problems:** A. The language often makes no sense. B. "Blah, blah, blah" is all that the reader reads and hears. C. Words are used incorrectly, making the message secondary to the misfires with the words. D. The lack of vocabulary and the misuse of parts of speech keep the reader from understanding. E. Repetition of words and phrases misuse of words and phrases litter the piece. F. Problems with language leave the reader wondering what the writer is trying to say. The words just don't work in this piece.

Sentence fluency

6 **The writing has an easy flow, rhythm and cadence. Sentences are well built, with strong and varied structure that invites expressive oral reading.** A. Sentences are constructed in a way that underscores and enhances the meaning. B. Sentences vary in length as well as structure. Fragments, if used, add style. Dialogue, if present, sounds natural. C. Purposeful and varied sentence beginnings add variety and energy. D. The use of creative and appropriate connectives between sentences and thoughts show how each relates to and builds upon the one before it. E. The writing has cadence; the writer has thought about the sound of the words as well as the meaning. The first time you read it aloud is a breeze.	**5** **Much of this piece has a sense of rhythm and flow, but some parts still need work. Technically the sentences are correctly structured.** A. Some of the sentences are phrased so carefully that the reader gets totally caught up in them; others remain a bit sterile. B. Correct construction is present in the sentences and variety in type is present. Few examples of risk-taking are present such as dialogue or fragments. C. Attention has been paid to different sentence beginners. Just a bit more attention here and the piece becomes musical. D. Connectives are present but not completely refined. E. You can read this piece aloud quite easily with only a moment or two of problems.	**4** **The text hums along with a steady beat, but tends to be more pleasant or businesslike than musical, more mechanical than fluid**. A. Although sentences may not seem artfully crafted or musical, they get the job done in a routine fashion. B. Sentences are usually constructed correctly; they hang together; they are sound. C. Sentence beginnings are not ALL alike; some variety is attempted. D. The reader sometimes has to hunt for clues (e.g., connecting words and phrases *like however*, *therefore*, *naturally*, *after a while*, *on the other hand, to be specific, for example, next, first of all, later, but as it turned out, although,* etc.) that show how sentences interrelate. E. Parts of the text invite expressive oral reading; others may be stiff, awkward, choppy, or gangly.
3 **Technically correct sentences tend to create a sing- song pattern or lull the reader to sleep**. **Nothing in the sentences creates a sense of fluidity.** A. Sentences are generally correct although a few may be lacking some key ingredients. B. You can read through the editing problems in this piece and see where the sentences logically begin and end. C. There is a reliance on patterned sentence beginnings; however, a few sentences break out. D. Only a very few and very simple connectives lead the reader from sentence to sentence. E. You can read this aloud - after a few tries.	**2** **Even some of the easier sentences have structural problems which cause the reader to stop and figure out what is being said and how.** A. The phrasing doesn't sound natural because of problems in structure as well as placement of words. B. To make the sentences correct and flow Many would have to be reconstructed. C. Many sentences begin the same way—and may follow the same patterns (e.g., subject-verb-object) in a monotonous pattern. D. Connectives, though present, are often misused or lead the reader in the wrong direction. E. The text does not invite expressive oral reading.	**1** **The reader has to practice quite a bit in order to give this paper a fair interpretive reading. The writing reflects more than one of the following problems:** A. Sentences are choppy, incomplete, rambling or awkward; they need work. B. There is little to no "sentence sense" present. Even if this piece was flawlessly edited, the sentences would not hang together. C. So many sentences are incomplete that it is hard to judge the quality of the beginnings. D. Endless connectives (*and, and so, but then, because, and then,* etc.) or a complete lack of connectives create a massive jumble of language. E. The text is so flawed that it cannot be read aloud without the writer's help.

Conventions

6 **The writer demonstrates a good grasp of standard writing conventions (e.g., spelling, punctuation, capitalization, grammar, usage, paragraphing) and uses conventions effectively to enhance readability. Errors tend to be so few that just minor touch-ups would get this piece ready to publish.** A. Spelling is generally correct, even on more difficult words. B. The punctuation is accurate, even creative, and guides the reader through the text. C. A thorough understanding and consistent application of capitalization skills are present. D. Paragraphing tends to be sound and reinforces the organizational structure. E. Grammar and usage are correct and contribute to clarity and style. F. The writer may manipulate conventions for stylistic effect - and it works! The piece is very close to being ready to publish. *Grades 7 & up only: The writing is sufficiently complex to allow the writer to show skill in using a wide range of conventions. For younger writers, the writing shows control over those conventions that are grade/age appropriate.*	**5** **The writer stretches and tries more complex tasks in conventions however makes a few mistakes along the way.** A. Everyday words are consistently handled well but more difficult words are spotty. B. Punctuation shows strength and enhances the readability in all but a few cases. C. The punctuation is usually correct and takes a few risks. D. Solid paragraphing skills are present although there may be a few adjustments needed on more complex pieces. E. The grammar and usage is correct. F. Just a few things here and there need to be edited before this piece is ready to publish.	**4** **The writer shows reasonable control over a limited range of standard writing conventions. Conventions are sometimes handled well and enhance readability; at other times, errors are distracting and impair readability.** A. Spelling is usually correct or reasonably phonetic on common words, but more difficult words are problematic. B. End punctuation is usually correct; internal punctuation (commas, apostrophes, semicolons, dashes, colons, parentheses) is sometimes missing/wrong. C. Most words are capitalized correctly; control over more sophisticated capitalization skills may be spotty. D. Paragraphing is attempted but may run together or begin in the wrong places. E. Problems with grammar or usage are not serious enough to distort meaning but may not be correct or accurately applied all of the time. F. Moderate (a little of this, a little of that) editing would be required to polish the text for publication.
3 **The writer stumbles in conventions even on simple tasks and almost always on anything trickier.** A. Although the reader can understand, even simpler words are not always correct. B. Punctuation is spotty and inconsistent. C. Proper nouns and the beginning of sentences are capitalized correctly, other words are random and don't show understanding of capitalization rules. D. The piece may start off with a paragraph or two, but then the rest is one big glob of sentences. E. There are Serious grammar and usage problems scattered throughout the text. F. Enough editing would have to be done to this piece that a student writer may need help to find it all.	**2** **Many errors of a variety of types are scattered throughout the text.** A. The spelling is phonetic, many errors are present. B. Except for the simplest of punctuation (periods, questions marks), the other punctuation is usually wrong or missing. C. Only the easiest rules of capitalization show awareness of correct use. D. Paragraphing skills are irregular and inconsistent. E. A heavy reliance on conversational oral language affects the grammar in an inappropriate way for this piece. F. Whew! There's quite a bit to be done here to edit the piece for publication.	**1** **Errors in spelling, punctuation, capitalization, usage and grammar and/or paragraphing repeatedly distract the reader and make the text difficult to read. The writing reflects more than one of these problems:** A. Spelling errors are frequent, even on common words. B. Punctuation (including terminal punctuation) is often missing or incorrect. C. Capitalization is random. D. Paragraphing is missing, irregular, or so frequent (every sentence) that it has no relationship to the organizational structure of the text. E. Errors in grammar or usage are very noticeable, frequent, and affect meaning. F. The reader must read once to decode, then again for meaning. Extensive editing (virtually every line) would be required to polish the text for publication.

Appendix F. Correlations between holistic and writing trait scores

This appendix presents the correlations between the holistic score and the writing trait scale scores at pretest (table F1) and posttest (table F2).

Table F1. Correlations between pretest holistic score and individual writing trait scores

Trait	Holistic	Ideas	Organization	Voice	Word choice	Sentence fluency
Holistic						
Ideas	0.65					
Organization	0.61	0.68				
Voice	0.56	0.71	0.60			
Word Choice	0.49	0.56	0.49	0.55		
Sentence Fluency	0.60	0.60	0.58	0.57	0.62	
Conventions	0.63	0.54	0.55	0.52	0.54	0.67

Source: Authors' analysis of 2009 and 2010 student writing assessment data.

Table F2. Correlations between posttest holistic score and individual writing trait scores

Trait	Holistic	Ideas	Organization	Voice	Word choice	Sentence fluency
Holistic						
Ideas	0.66					
Organization	0.62	0.72				
Voice	0.58	0.75	0.64			
Word Choice	0.55	0.62	0.58	0.61		
Sentence Fluency	0.63	0.66	0.63	0.65	0.70	
Conventions	0.64	0.60	0.61	0.58	0.62	0.72

Source: Authors' analysis of 2009 and 2010 student writing assessment data.

Appendix G. Interrater reliability and coding details for student essays

The procedures used to score the student essays in this research study were similar to those used by the Oregon Department of Education to score grade 4 and grade 7 statewide writing assessments. The quality control procedures also aligned with accepted practice for statewide writing assessment in Oregon (Oregon Department of Education 2005).

Student essay scoring process and quality control measures

Rating team members, qualifications, and training. The assessment team that scored the student essays for this study included an assessment coordinator and 11 writing assessment raters who were well trained and experienced in procedures for scoring K–12 student writing essays.

Assessment coordinator. The assessment coordinator for this project had more than two decades of experience assessing student writing and 10 years of experience training educators to score student writing samples using both analytic and holistic rubrics for school-age students. For this study, the coordinator was responsible for training the raters in the appropriate scoring rubrics, organizing the student essay rating assignments, and continuously monitoring the scoring quality and consistency across raters.

Raters. Eleven raters formed the teams that completed the holistic and analytic scoring for both cohorts of student essays. All raters were or had been certified classroom teachers and were considered well qualified and trained in writing assessment. The raters selected for this research project had previous training and experience using the 6+1 Trait analytic scoring rubric and other rubrics to score student essays.

Initial training for raters. Before scoring each cohort of student essays, the coordinator conducted a four- to six-hour training session with all members of the rating team. The initial training included a thorough review of the analytic or holistic scoring rubric, the definitions of the special codes used to document the reason an essay was not scored, and procedures for entering student scores using the online data system. The raters also practiced scoring sample grade 5 essays, both as a group and individually, using the appropriate scoring rubric. If the raters' scores for a given essay were different, the coordinator used the language of the rubric to discuss the rationale for converging on a particular score. The coordinator continued the initial training session until the raters demonstrated consistent scoring on practice papers.

The coordinator conducted a review of the rubrics and scoring procedures each Monday of the scoring session. During these sessions, raters reviewed the rubric and scored practice essays to ensure that consistency in the process and criteria for scoring essays

was maintained. The coordinator also provided additional coaching and individual training for any rater that demonstrated interrater agreement scores of 0.95 or less.

Student essay scoring procedures. All completed student essays were sent directly to Chesapeake Research Associates, an external research agency, by the participating schools. For each cohort, Chesapeake Research Associates assigned new unique identification numbers to the essays and randomly mixed the pretests and posttests together. It then sent all the essays to the Regional Educational Laboratory (REL) Northwest assessment team for scoring. This procedure ensured that the raters were blind to whether the essay was a pretest or posttest and whether it was from a treatment or control group school.

The scoring process for each cohort of student essays was conducted during two separate sessions scheduled during different months. In one session, raters used the six analytic dimensions from the 6+1 Trait analytic rubric to score the student essays; in the other session, raters used the holistic scoring rubric. The assessment coordinator assigned two different raters to score each student essay. The raters were blind to the identity of the second rater and were placed in different locations to score the essays. Raters were instructed not to discuss the essays or scoring questions with each other. All rater questions were directed to the assessment coordinator to ensure that communication about scoring was consistent across the rating team. The coordinator also changed the pairing of raters throughout the scoring session as an additional strategy for identifying scoring inconsistencies among the raters.

Scoring decisions. Both the analytic and holistic rubric used a six-point scale on which 1 indicated a writing sample that demonstrated very little to no mastery and 6 indicated a writing sample that met or exceeded the highest standard. If the scores of the two raters were identical or one point part, the final score assigned to the essay was calculated as an average of the two raters. For example, if a student essay was assigned a 2 by one rater and a 3 by a second rater, the final essay score was 2.5.

If the scores of the two raters were more than one point apart, the essay was flagged as discrepant and scored by a third rater, the assessment coordinator. The flagging system was designed to alert the coordinator that a student essay required a third read without disclosing the scores already assigned by the original two raters. In these cases, the assessment coordinator then scored the essay and asked the two original raters to read the essay again and revise their scores in consultation with the coordinator. This procedure was used for continuous quality improvement of the rating team and as a quality assurance procedure for the individual essay scores; it is a common practice to improve the reliability of ratings of open-ended assessments (Johnson et al. 2005).

For the holistic scores, a total of 50 essays (12 treatment group essays and 8 control group essays at pretest, 15 treatment group essays and 15 control group essays at posttest) required intervention by the assessment coordinator (0.6 percent of all scored essays). For the individual trait scores, 177 essays (48 treatment group essays and 44 control group essays at pretest, 45 treatment group essays and 40 control group essays at posttest)

required intervention by the assessment coordinator (2.2 percent of all scored essays). The raters applied standard rules for flagging essays with special codes indicating reasons they could not be scored; these codes were built into the data management system used by Education Northwest for all writing assessment projects. If the rater determined that a student essay could not be scored, he or she documented the reason for this decision using a special code (table G1).

Of the 4,283 returned student pretest essays, 169 were not scored by raters because they were blank, too short to score, written in a language other than English, or illegible. Another 208 essays were coded as being off topic. Off topic essays were scored by the raters using the regular scoring procedures. Two records were missing special codes or scoring data.

Of the 3,932 returned student posttest essays, 205 were not scored by raters because they were blank, too short to score, written in a language other than English, or illegible. Another 108 essays were coded as off topic.

Table G1. Number and percentage of unrated student essays, by reason

	Treatment		Control		Total	
Reason	Number	Percent	Number	Percent	Number	Percent
Pretest						
No special code	2,126	91.1	1,778	91.2	3,904	91.2
Special codes						
Blank or missing data	61	2.6	67	3.4	128	3.0
Too short, not in English, or illegible	24	1.0	19	1.0	43	1.0
Scored but off topic	122	5.2	86	4.4	208	4.9
Total	2,333	100	1,950	100	4,283	100
Posttest						
No special code	1,949	92.2	1,670	91.9	3,619	92.0
Special codes						
Blank or missing data	91	4.3	96	5.3	187	4.8
Too short, not in English, or illegible	10	0.5	8	0.4	18	0.5
Scored but off topic	64	3.0	44	2.4	108	2.7
Total	2,114	100	1,818	100	3,932	100

Note: Percentages may not sum to 100 because of rounding.
Source: Authors' analysis of 2009 and 2010 student writing assessment data.

Online writing assessment data system

Education Northwest maintains an online, browser-based software system to manage large-scale writing assessment projects. In addition to managing student writing assessment scores, the online system provides a variety of reporting features—such as reporting the number of discrepancies, real-time interrater agreement statistics for each rater, and project summary reports—that are used to continuously monitor scoring quality and consistency. Access to student essay data stored in the online writing assessment system is restricted to staff members directly involved in scoring student essays (the assessment coordinator, the online system administrator, and the raters). Permissions to access specific scoring data and reports are restricted based on three authorization levels: rater access, coordinator access, and online system administrator access. This preexisting data management system was used for this study.

Rater access. Raters used the online system to enter assessment data for each student. Rater access to the system was restricted to obtaining the data entry screen that corresponded to the student essay code number, online recording of assessment scores, and, if necessary, special code information. Indicator flags associated with each student record allowed the rater to determine if the student essay required scoring, had been scored once, or had been scored twice. Raters did not have access to the identity of the other raters that were assigned to, or had completed scoring of, any student essay. Raters also did not have access to the scores assigned to an essay by another rater.

Coordinator access. The assessment coordinator had access to the scoring information entered by the raters and the following real-time project monitoring reports: the number of essays read daily by each rater and for the project as a whole, the identification numbers of essays that were discrepant and required a third read, and the number of times raters changed their original scores. The data management system also provides summary tables that report the frequency of discrepant scores and interrater agreement statistics for each rater.

Administrator access. The online system administrator was the only person allowed to change the data entry, storage, and quality control check functions for the online system. The system administrator was the only person who had permissions to access all student record numbers, scoring data, and rater information stored in the data system.

Interrater reliability among the rating teams. The assessment coordinator randomly selected a set of student essays for all raters to score before the 2009 analytic scoring session and scoring of both cohorts using the holistic rubric. The common set of essay was not scored by raters scoring the second cohort using the analytic rubrics. All raters participating in the project produced independent scores for these papers. Intraclass correlation coefficients (ICCs) were then calculated to determine the extent to which variance in the matrix of student scores was related to the essays themselves rather than rater differences. The use of ICC estimates to calculate interrater reliability is considered a more accurate method than the more common correlation statistics that compare paired observations (Bartko 1991; McGraw and Wong 1996; Shrout and Fleiss 1979). The ICC

results for the raters who scored the 2009 cohort using analytic rubrics are presented in table G2; the ICC results for the rating teams that used the holistic rubrics are presented in table G3. In this context, the ICC values indicate the proportion (percentage) of total variance in each trait score that is related to student essay variation; the proportion of variance in the scores that is related to rater differences is 1 minus the ICC. All reliability coefficients were at or above 0.94, meaning that 94 percent or more of the variation in scores was due to differences in the essays rather than differences among raters.

Table G2. Variance estimates and intraclass correlation coefficient results for analytic scoring of the "common" set of student essays

Trait	Rater	Total	Intraclass correlation coefficient
Ideas	0.97	51.97	0.981
Organization	1.97	43.97	0.955
Voice	1.90	34.00	0.944
Word choice	0.88	34.93	0.975
Sentence fluency	0.64	48.59	0.987
Conventions	1.68	61.13	0.973

Note: Number of raters was 7; number of essays was 20.
Data are from the first cohort of students only.
a. Intraclass correlation coefficient = [1–rater variance]/total variance.
Source: Authors' analysis of 2009 student writing assessment data.

Table G3. Variance estimates and intraclass correlation coefficient results for holistic scoring of the "common" set of student essays

	Variance		Intraclass correlation coefficient	Number of raters	Number of essays
Cohort	Rater	Total			
2009	4.54	87.74	0.948	7	20
2010	0.93	22.93	0.959	4	15

a. Intraclass correlation coefficient = [1–rater variance]/total variance.
Source: Authors' analysis of 2009 and 2010 student writing assessment data.

Appendix H. Technical note on multiple imputation of missing data

Students were included in the analysis only if they were present at the baseline measure and met eligibility requirements. These criteria resulted in 4,161 students in the analysis, including 2,230 students in treatment schools and 1,931 in control schools. In the treatment schools, 116 students (5.2 percent) had a missing posttest score and were therefore classified as "leavers." In the control schools, 113 students (5.9 percent) had a missing posttest score and were classified as leavers. Leavers were included in the analysis to arrive at an intent-to-treat (ITT) estimate of the treatment effect.

Analysis of "stayers" versus "leavers" revealed that leavers were likely to have lower pretest scores than stayers, to have attended schools with higher percentages of students receiving free or reduced-price lunches (*FRL%*), to be members of minority groups, and to be boys. For students in the treatment schools, the mean pretest score was 3.61 (95 percent confidence interval [CI]: 3.58–3.64) for stayers and 3.38 (CI: 3.24–3.52) for leavers. For students in the control schools, the mean pretest score was 3.69 (CI: 3.66–3.73) for stayers and 3.35 (CI: 3.17–3.53) for leavers.

Stayers and leavers in treatment schools did not differ with respect to *FRL%* or ethnicity. In contrast, in control schools, the mean *FRL%* was 45.9 percent (CI: 45.0–46.8 percent) for stayers and 54.8 percent (CI: 51.0–58.5 percent) for leavers. Leavers were more likely to be boys in both treatment and control schools.

Ethnicity information was missing on only 3 students (0.1 percent) at treatment schools and 7 students (0.4 percent) at control schools. Another 17 students (0.8 percent) at treatment schools and 49 students (2.5 percent) at control schools marked their ethnicity as "unknown." Gender information was missing on five students (0.2 percent) at the treatment schools and no student at the control schools.

The rate of missing outcome data exceeded the preset cutoff of 5 percent, below which the researchers planned to use listwise deletion. As a result, multiple imputation (MI) was used to handle missing data.

Some students received no pretest score even though they were present at the pretest and turned in pretest booklets. Examination of these cases revealed that 169 students (84 control, 85 treatment) turned in essays that were too short to score or were blank. These students were assigned the lowest possible pretest score. A similar process resulted in 205 (104 control, 101 treatment) low posttest score assignments for students who were present but had not received a score from the rating team.

Stata's MI IMPUTE command was then used to impute remaining missing data on the posttreatment assessment as well as on gender. As a set of nonmissing variables, the design variable School ID was included in the multiple imputation. Ethnicity, which was originally coded as a multinomial categorical variable, was recoded using a set of dummy

variables. Missing responses to ethnicity were coded together with those designated as "unknown," eliminating missing data in the ethnicity variables. Ethnicity was then added as another nonmissing variable for imputation. Because multiple variables have missing data, and no sequential pattern of missingness was observed, the imputation was done with the assumption of multivariate normality. Five sets of complete data were imputed, separately for the treatment and control groups. Those 10 imputed datasets were then merged for the impact analysis. The imputation was done using an iterative procedure based on the Markov chain Monte Carlo method. Proper convergence behavior was ensured using a trace plot and an autocorrelation plot for the worst linear function.

Appendix I. Alternative analysis of implementation fidelity

The teacher survey was administered to both treatment and control group teachers to assess their use of the classroom practices emphasized in the 6+1 Trait Writing intervention. Although most of the questions did not ask specifically about the use of writing traits, 13 of the 45 teacher survey questions referred to the use of writing traits in general, without referencing details of the 6+1 Trait intervention. The survey scores were recalculated without those items in order to determine whether the exclusion of items that referred to writing traits affected the implementation fidelity findings.

Excluding these items had no effect on the survey results at baseline (table I1) except that the scale for "teaching the language of rubrics" no longer existed because all items on that scale referred to writing traits. At midyear (table I2), differences between treatment and control groups on two scales ("teaching focused revision strategies" and "modeling participation in the writing process") were no longer significant after dropping items that referred to writing traits. One scale ("giving students writing assignments to respond to effective prompts") had previously shown no significant difference and remained nonsignificant in the alternative analysis. Teachers in the treatment group reported higher levels of use of six of the nine strategies that remained after the loss of the "teaching the language of rubrics" scale, and had significantly higher total survey scores than did the control group teachers. At the end of the year, the exclusion of these items had no effect on the survey results (table I3) except that the scale for "teaching the language of rubrics" no longer existed. Differences between treatment and control group teachers remained significant on all nine remaining scales.

Table I1. Baseline teacher-reported use of instructional strategies for writing, excluding questions referring to writing traits

Instructional strategy (number of survey items contributing to the scale)	Coefficient alpha	Mean scale score[a]		Difference	Test statistic
		Treatment group	Control group		
Reading and scoring papers and justifying the scores (5)	0.79	2.34	2.56	–0.22	$t = –1.25$
		(1.15)	(1.16)		$p = .212$
Teaching focused revision strategies (2)	0.74	4.13	4.36	–0.23	$t = –1.35$
		(1.20)	(1.20)		$p = .180$
Modeling participation in the writing process (2)	0.73	2.80	2.73	0.07	$t = 0.33$
		(1.46)	(1.52)		$p = .739$
Having students read and analyze materials that demonstrate varying writing quality (2)	0.75	2.50	2.64	–0.14	$t = –0.72$
		(1.35)	(1.33)		$p = .472$
Giving students writing assignments to respond to effective prompts (4)	0.83	3.60	3.60	–0.00	$t = –0.01$
		(1.07)	(1.12)		$p = .993$
Weaving writing lessons into other subjects (4)	0.77	3.11	3.29	–0.18	$t = –1.09$
		(1.13)	(1.12)		$p = .278$
Teaching students to set goals and monitor their progress (5)	0.78	2.85	2.88	–0.03	$t = –0.18$
		(1.16)	(1.19)		$p = .860$
Integrating learning goals for writing into curriculum planning (4)	0.78	3.81	3.76	0.05	$t = 0.35$
		(0.96)	(1.04)		$p = .728$
Teaching ways to structure nonfiction writing (4)	0.80	2.99	3.19	–0.20	$t = –1.11$
		(1.23)	(1.16)		$p = .268$
Total score	0.92	3.13	3.22	–0.09	$t = –0.66$
		(0.94)	(0.94)		$p = .510$

Note: Total score is the mean of the nine scale scores after dropping items referencing writing traits. Numbers in parentheses are standard deviations.
a. Scores ranged from zero to six.
Source: Authors' analysis, based on data described in text.

Table I2. Midyear teacher-reported use of instructional strategies for writing, excluding questions referring to writing traits

Instructional strategy (number of survey items contributing to the scale)	Coefficient alpha	Mean scale score[a]		Difference	Test statistic
		Treatment group	Control group		
Reading and scoring papers and justifying the scores (5)	0.82	3.49	2.58	0.91	$t = 5.41$
		(1.07)	(1.22)		$p \leq .001$
Teaching focused revision strategies (2)	0.76	4.34	4.29	0.05	$t = 0.34$
		(0.96)	(1.16)		$p = .731$
Modeling participation in the writing process (2)	0.70	3.10	2.74	0.36	$t = 1.70$
		1.42	(1.46)		$p = .093$
Having students read and analyze materials that demonstrate varying writing quality (2)	0.60	3.48	2.76	0.72	$t = 4.14$
		(1.08)	(1.27)		$p \leq .001$
Giving students writing assignments to respond to effective prompts (4)	0.80	3.87	3.60	0.27	$t = 1.72$
		(1.03)	(1.09)		$p = .088$
Weaving writing lessons into other subjects (4)	0.76	3.55	3.16	0.39	$t = 2.50$
		(1.06)	(1.08)		$p = .013$
Teaching students to set goals and monitor their progress (5)	0.83	3.34	2.85	0.49	$t = 2.76$
		(1.02)	(1.34)		$p = .006$
Integrating learning goals for writing into curriculum planning (4)	0.81	4.21	3.76	0.45	$t = 3.19$
		(0.91)	(0.99)		$p = .002$
Teaching ways to structure nonfiction writing (4)	0.78	3.47	3.10	0.37	$t = 2.33$
		(1.08)	(1.05)		$p = .021$
Total score	0.92	3.65	3.20	0.45	$t = 3.46$
		(0.82)	(0.95)		$p = .001$

Note: Total score is the mean of the nine scale scores after dropping items referencing writing traits. Numbers in parentheses are standard deviations.

a. Scores ranged from zero to six.

Source: Authors' analysis, based on data described in text.

Table I3. End-of-year teacher-reported use of instructional strategies for writing, excluding questions referring to writing traits

Instructional strategy (number of survey items contributing to the scale)	Coefficient alpha	Mean scale score[a]		Difference	Test statistic
		Treatment group	Control group		
Reading and scoring papers and justifying the scores (5)	0.85	3.86	2.74	1.12	$t = 6.57$
		(1.05)	(1.22)		$p \leq .001$
Teaching focused revision strategies (2)	0.77	4.67	4.29	0.38	$t = 2.49$
		(0.87)	(1.16)		$p = .014$
Modeling participation in the writing process (2)	0.82	3.39	2.80	0.59	$t = 2.71$
		(1.43)	(1.51)		$p = .007$
Having students read and analyze materials that demonstrate varying writing quality (2)	0.70	3.82	2.86	0.96	$t = 5.52$
		(1.12)	(1.22)		$p \leq .001$
Giving students writing assignments to respond to effective prompts (4)	0.82	4.32	3.77	0.55	$t = 3.54$
		(0.94)	(1.13)		$p = .001$
Weaving writing lessons into other subjects (4)	0.81	4.05	3.36	0.69	$t = 4.40$
		(0.95)	(1.17)		$p \leq .001$
Teaching students to set goals and monitor their progress (5)	0.85	3.74	2.99	0.75	$t = 4.24$
		(1.10)	(1.29)		$p \leq .001$
Integrating learning goals for writing into curriculum planning (4)	0.84	4.40	3.89	0.51	$t = 3.55$
		(0.86)	(1.06)		$p \leq .001$
Teaching ways to structure nonfiction writing (4)	0.82	3.93	3.37	0.56	$t = 3.46$
		(1.06)	(1.11)		$p = .001$
Total score	0.94	4.02	3.34	0.68	$t = 4.98$
		(0.84)	(1.00)		$p \leq .001$

Note: Total score is the mean of the nine scale scores after dropping items referencing writing traits. Numbers in parentheses are standard deviations.

a. Scores ranged from zero to six.

Source: Authors' analysis, based on data described in text.

Appendix J. Complete multilevel model results for confirmatory research question

The model parameter estimates used the five multiply imputed datasets (MI sets). Each MI set was first split into two segments: one belonging to paired schools and the other belonging to singleton schools. The parameter estimations were performed separately for paired schools and for singleton schools, using the appropriate segments. Both the estimates for paired schools and those for singleton schools were based on the pooling of five sets of individual parameter estimates, each arrived at by applying a linear mixed model (LMM) to one of the five multiply imputed datasets. LMMs are based on maximum likelihood estimation (MLE). Table J1 presents the parameter estimates for paired schools; whereas J2 presents those for singleton schools. Those separate estimates were later pooled using the method of inverse-variance weighting (appendix M).

Table J1: Estimation for pairs

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
trt	.1297651	.0455003	2.85	0.004	1317.1
gmctr_PRE_~b	.4071435	.0189835	21.45	0.000	1487.5
q47	-.0078661	.009738	-0.81	0.419	2668.3
q52	-.0375622	.0161408	-2.33	0.020	2196.7
q53	.0446303	.0197124	2.26	0.024	1626.9
_cons	3.612587	.1238489	29.17	0.000	48678.0
/sigma_u	.0893119	.0213748			3044.7
/sigma_e	.8091247	.0101348			10629.9
rho	.0120373	.0057292			

Table J2: Estimation for singletons

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
trt	-.2833468	.1620851	-1.75	0.080	98087.3
gmctr_PRE_~b	.4166048	.0400958	10.39	0.000	2209.6
q47	.0508619	.0444474	1.14	0.253	22958.7
q52	-.0190521	.0231597	-0.82	0.411	5126.4
q53	.0095358	.0230608	0.41	0.679	29233.2
frl	-.0013948	.0030596	-0.46	0.649	2414.4
_cons	4.131015	.2628803	15.71	0.000	639.5
/sigma_u	.1446079	.0451445			471.7
/sigma_e	.7947321	.0203034			1377.3
rho	.0320477	.0195848			

Note: trt = treatment effect; gmctr_PRE~b = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced lunch; cons = intercept; sigma_u = school level random effect; sigma_e = student level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Appendix K. Technical note on effect size calculations

The treatment effect on the original scale is difficult to interpret unless one is familiar with the measurement instrument. Consequently, this value was standardized using the control condition standard deviation of the posttest score (Glass's delta).

Appendix L. Complete multilevel model results for exploratory analyses

The model parameter estimates used the five multiply imputed datasets (MI sets). Each MI set was first split into two segments: one belonging to paired schools, and the other belonging to singleton schools. The parameter estimations were performed separately for paired schools and for singleton schools, using the appropriate segments. Both the estimates for paired schools and those for singleton schools were based on the pooling of five sets of individual parameter estimates, each arrived at by applying a linear mixed model (LMM) to one of the five multiply imputed datasets. LMMs are based on maximum likelihood estimation (MLE). Tables are presented as pairs. The first of the pair, with a suffix "a" presents the results for paired schools. The second of the pair, with a suffix "b" presents those for singleton schools. Those separate estimates were later pooled using the method of inverse-variance weighting (appendix M).

Model results for Ideas trait scale

Table L1a: Estimation for pairs: Ideas

POST_score_I	Coef.	Std. Err	t	P> | t |	DF
trt	.1064809	.0410149	2.60	0.013	39.0
gmctr_PRE_~I	.3833639	.0216252	17.73	0.000	45.4
q47	-.0207579	.0079217	-2.62	0.010	178.0
q52	-.0355988	.0139497	-2.55	0.013	63.6
q53	.0491644	.0175057	2.81	0.007	45.8
_cons	4.069582	.0991918	41.03	0.000	407.8
/sigma_u	.0689926	.016503			380.8
/sigma_e	.6133121	.013985			8.1
rho	.0124963	.0059432			

Table L1b: Estimation for singletons: Ideas

POST_score_I	Coef.	Std. Err	t	P> | t |	DF
trt	-.3395874	.0970959	-3.50	0.000	2766.8
gmctr_PRE_~I	.396038	.041242	9.60	0.000	171.2
q47	.0214114	.028063	0.76	0.446	532.6
q52	-.0475871	.0137688	-3.46	0.001	958.6
q53	.0336216	.0138921	2.42	0.016	1217.3
frl	-.002312	.0018069	-1.28	0.201	9632.0
_cons	4.802456	.1594099	30.13	0.000	321.1
/sigma_u	.0662275	.0364779			124.4
/sigma_e	.6108876	.0167321			128.4
rho	.0116166	.0127015			

Note: trt = treatment effect; gmctr_PRE~I = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results for Organization trait scale

Table L2a: Estimation for pairs: Organization

POST_score_O	Coef.	Std. Err	t	P> | t |	DF
trt	.0804674	.0285917	2.81	0.007	51.8
gmctr_PRE~O	.3582155	.0206884	17.31	0.000	54.9
q47	-.017232	.0057581	-2.99	0.004	79.7
q52	-.0176769	.0094512	-1.87	0.063	150.4
q53	.0216674	.0114762	1.89	0.061	162.4
_cons	3.872955	.0702757	55.11	0.000	16035.9
/sigma_u	.0174424	.0319274			202.2
/sigma_e	.5555827	.0099518			14.7
rho	.0009847	.0036083			

Table L2b: Estimation for singletons: Organization

POST_score_O	Coef.	Std. Err	t	P> | t |	DF
trt	-.2789917	.1155339	-2.41	0.016	2599.0
gmctr_PRE~O	.3806634	.040693	9.35	0.000	156.3
q47	.0229405	.0313047	0.73	0.464	12247.3
q52	-.023169	.0166656	-1.39	0.165	551.0
q53	.0099609	.0166679	0.60	0.550	622.2
frl	-.0013152	.0021351	-0.62	0.538	3588.0
_cons	4.335343	.1843155	23.52	0.000	596.9
/sigma_u	.1006102	.0327585			163.3
/sigma_e	.5593199	.0201436			14.6
rho	.0313425	.0195464			

Note: trt = treatment effect; gmctr_PRE~O = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results for Voice trait scale

Table L3a: Estimation for pairs: Voice

POST_score_V	Coef.	Std. Err	t	P> | t |	DF
trt	.0922743	.0283736	3.25	0.001	155.0
gmctr_PRE~V	.3684706	.0191743	19.22	0.000	467.7
q47	-.0094016	.006067	-1.55	0.123	193.8
q52	-.0234138	.0101178	-2.31	0.022	160.3
q53	.0318667	.0126039	2.53	0.013	103.7
_cons	4.311279	.0737009	58.50	0.000	3982.2
/sigma_u	.0517802	.0139122			184.2
/sigma_e	.4789776	.0072			39.4
rho	.0115518	.0061709			

Table L3b: Estimation for singletons: Voice

POST_score_V	Coef.	Std. Err	t	P> | t |	DF
trt	-.3324087	.1032184	-3.22	0.001	30114.7
gmctr_PRE~V	.34346	.0386131	8.89	0.000	1498.0
q47	.0482833	.0281704	1.71	0.087	12040.0
q52	-.0370115	.0147734	-2.51	0.012	2164.7
q53	.0259748	.0146546	1.77	0.076	6571.5
frl	-.0015252	.0019145	-0.80	0.426	11745.4
_cons	4.761306	.1636189	29.10	0.000	1458.2
/sigma_u	.0950192	.0272032			652.8
/sigma_e	.4685805	.0153806			22.0
rho	.0394961	.0217694			

Note: trt = treatment effect; gmctr_PRE~V = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results for Word Choice trait scale

Table L4a: Estimation for pairs: Word Choice

POST_score_W	Coef.	Std. Err	t	P> | t |	DF
trt	.0860313	.0249024	3.45	0.001	47.6
gmctr_PRE~W	.4192989	.024557	17.07	0.000	173.7
q47	-.0124996	.0048657	-2.57	0.011	202.0
q52	-.0301563	.0080397	-3.75	0.000	232.7
q53	.0372106	.0101544	3.66	0.000	107.4
_cons	3.982832	.0656733	60.65	0.000	99.0
/sigma_u	.0400962	.0111154			378.3
/sigma_e	.3965546	.0070837			14.9
rho	.0101201	.0055426			

Table L4b: Estimation for singletons: Word Choice

POST_score_W	Coef.	Std. Err	t	P> | t |	DF
trt	-.1704965	.0675285	-2.52	0.012	2060.0
gmctr_PRE~W	.4417344	.0480517	9.19	0.000	3360.7
q47	.0323177	.0187365	1.72	0.085	1581.5
q52	-.0145962	.009757	-1.50	0.135	429.2
q53	.008064	.0099481	0.81	0.418	292.5
frl	-.0010439	.0012921	-0.81	0.420	457.3
_cons	4.297689	.1047877	41.01	0.000	24887.9
/sigma_u	.0506818	.0257769			33.6
/sigma_e	.3843698	.0102695			221.4
rho	.0170891	.0171414			

Note: trt = treatment effect; gmctr_PRE~W = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results for Sentence Fluency trait scale

Table L5a: Estimation for pairs: Sentence Fluency

POST_score_S	Coef.	Std. Err	t	P> | t |	DF
trt	.0914604	.0316418	2.89	0.006	48.6
gmctr_PRE~S	.4034601	.0192435	20.97	0.000	250.1
q47	-.0102315	.0059689	-1.71	0.087	758.7
q52	-.0283322	.0105094	-2.70	0.008	129.9
q53	.0347711	.0130633	2.66	0.009	90.6
_cons	3.917338	.0798161	49.08	0.000	248.7
/sigma_u	.0523783	.013114			5145.8
/sigma_e	.4977062	.0099834			10.5
rho	.010954	.0054245			

Table L5b: Estimation for singletons: Sentence Fluency

POST_score_S	Coef.	Std. Err	t	P> | t |	DF
trt	-.2467081	.0877538	-2.81	0.005	740.1
gmctr_PRE~S	.4401844	.0449277	9.80	0.000	56.6
q47	.0286516	.0244538	1.17	0.242	510.8
q52	-.0141686	.0127228	-1.11	0.267	240.7
q53	.0042272	.0128263	0.33	0.742	234.8
frl	-.001263	.0016237	-0.78	0.437	1058.2
_cons	4.384639	.1392939	31.48	0.000	447.4
/sigma_u	.0686539	.0254819			333.1
/sigma_e	.4763059	.0188063			11.0
rho	.020353	.014653			

Note: trt = treatment effect; gmctr_PRE~S = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results for Conventions trait scale

Table L6a: Estimation for pairs: Conventions

POST_score_C	Coef.	Std. Err	t	P> | t |	DF
trt	.0655367	.0350996	1.87	0.065	95.0
gmctr_PRE~C	.5062013	.0209887	24.12	0.000	42.5
q47	-.0046368	.0069012	-0.67	0.502	5884.8
q52	-.0249199	.0119738	-2.08	0.038	246.6
q53	.028358	.0146756	1.93	0.055	203.4
_cons	3.9143	.0915871	42.74	0.000	348.7
/sigma_u	.0701877	.0138628			589.4
/sigma_e	.5258807	.0089958			17.8
rho	.0175017	.0068303			

Table L6b: Estimation for singletons: Conventions

POST_score_C	Coef.	Std. Err	t	P> | t |	DF
trt	-.2141265	.0895004	-2.39	0.017	729.6
gmctr_PRE~C	.5636923	.0385935	14.61	0.000	328.7
q47	.0369349	.0240829	1.53	0.125	16137.3
q52	-.0125587	.0124556	-1.01	0.314	1113.2
q53	.0034122	.0125957	0.27	0.787	962.3
frl	-.001217	.0016463	-0.74	0.460	2370.4
_cons	4.27601	.1395137	30.65	0.000	1518.8
/sigma_u	.0666261	.0263901			1694.8
/sigma_e	.5126145	.0148227			58.3
rho	.0166124	.0130154			

Note: trt = treatment effect; gmctr_PRE~C = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results comparing girls with boys

Table L7a: Estimation for pairs: girls compared with boys

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
1. trt	.0836895	.0529383	1.58	0.114	1693.4
1. male	-.2324299	.0426922	-5.44	0.000	383.3
trt#male 1 1	.0832414	.0584767	1.42	0.155	802.7
gmctr_PRE~b	.3827616	.0190364	20.11	0.000	3582.1
q47	-.0085387	.0095155	-0.90	0.370	2382.0
q52	-.0395574	.0157953	-2.50	0.012	2043.5
q53	.0456357	.0192879	2.37	0.018	1529.2
cons	3.733988	.1244278	30.01	0.000	7003.3
/sigma_u	.0851796	.0214217			4000.8
/sigma_e	.8037512	.0100918			6930.5
rho	.0111065	.0055591			

Table L7b: Estimation for singletons: girls compared with boys

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
1. trt	-.3522058	.1748141	-2.01	0.044	33752.5
1. male	-.3165519	.0991091	-3.19	0.001	71862.3
trt#male 1 1	.1508367	.1197622	1.26	0.208	11351.2
gmctr_PRE~b	.3864653	.0406891	9.50	0.000	1453.6
q47	.0549477	.0445632	1.23	0.218	25663.9
q52	-.0176013	.0232718	-0.76	0.449	5003.7
q53	.0086792	.0231651	0.37	0.708	24694.4
frl	-.0015232	.0030724	-0.50	0.620	2266.3
cons	4.266732	.2685033	15.89	0.000	637.6
/sigma_u	.1462104	.0449158			498.7
/sigma_e	.7869187	.0202084			987.0
rho	.0333701	.0200419			

Note: 1.trt – treatment effect; 1.male = gender main effect; trt#male 11 = moderator effect of gender; gmctr_PRE~b = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results for comparing White non-Hispanics with all others

Table L8a: Estimation for pairs: White non-Hispanics compared with all others

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
1. trt	.1208624	.0480628	2.51	0.012	3239.0
0.white	-.0521339	.0480391	-1.09	0.278	42539.6
trt#white 1 0	.0312417	.0692321	0.45	0.652	4892.6
gmctr_PRE)_~b	.4051387	.0190916	21.22	0.000	1372.8
q47	-.0083777	.0098246	-0.85	0.394	2463.2
q52	-.0384169	.0162729	-2.36	0.018	2114.9
q53	.0452496	.0198484	2.28	0.023	1538.8
cons	3.626445	.1251021	28.99	0.000	49697.1
/sigma_u	.0903878	.0214238			4437.8
/sigma_e	.8088795	.0101349			10271.5
rho	.0123328	.0058121			

Table L8b: Estimation for singletons: White non-Hispanics compared with all others

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
1. trt	-.2474389	.1708368	-1.45	0.148	112258.5
0.white	.1375439	.202343	0.68	0.497	455.4
trt#white 1 0	-.301666	.2218647	-1.36	0.175	394.0
gmctr_PRE~b	.414142	.0400254	10.35	0.000	2133.0
q47	.0472455	.0465876	1.01	0.311	17216.6
q52	-.0213945	.024298	-0.88	0.379	7648.6
q53	.0122629	.0242468	0.51	0.613	29161.2
frl	-.0010572	.0032013	-0.33	0.741	3134.1
cons	4.130631	.2755321	14.99	0.000	628.3
/sigma_u	.1556829	.0460749			523.6
/sigma_e	.791922	.0202846			1156.9
rho	.0372091	.0214576			

Note: 1.trt = treatment effect; 0.white = minority (non-White) main effect; trt#white 10 = moderator effect of ethnicity; gmctr_PRE~b = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Model results for comparing White non-Hispanics with Hispanics

Table L9a: Estimation for pairs: White non-Hispanics compared with Hispanics

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
1. trt	.1200826	.0474227	2.53	0.011	2425.3
1.hisp	-.1034084	.0700755	-1.48	0.141	250.1
trt#hisp 1 1	.0566957	.0931836	0.61	0.543	516.6
gmctr_PRE~b	.4016866	.0214284	18.75	0.000	268.9
q47	-.008194	.0100743	-0.81	0.416	1746.9
q52	-.0351235	.0172936	-2.03	0.042	1756.5
q53	.0409115	.020517	1.99	0.046	1974.0
cons	3.676425	.1259508	29.19	0.000	120209.3
/sigma_u	.0834371	.0235384			9079.8
/sigma_e	.8072473	.0108081			10567.9
rho	.0105704	.0059389			

Table L9b: Estimation for singletons: White non-Hispanics compared with Hispanics

POST~e_Hol_b	Coef.	Std. Err	t	P> | t |	DF
1. trt	-.2516066	.183061	-1.37	0.169	30686.0
1.hisp	.0999178	.2771553	0.36	0.719	166.4
trt#hisp 1 1	-.2903606	.299824	-0.97	0.334	218.0
gmctr_PRE~b	.4149222	.0416758	9.96	0.000	1430.6
q47	.0488794	.0502651	0.97	0.331	25343.0
q52	-.0230991	.025954	-0.89	0.373	6460.7
q53	.0149889	.0257959	0.58	0.561	32026.4
frl	-.000303	.0034086	-0.09	0.929	4395.8
cons	4.093467	.2911449	14.06	0.000	738.2
/sigma_u	.1692612	.0491343			457.9
/sigma_e	.7779621	.0210673			834.6
rho	.0451972	.0254135			

Note: 1.trt = treatment effect; 1.hisp = Hispanic main effect; trt#hisp 11 = moderator effect of Hispanic; gmctr_PRE~b = grand-mean centered prescore; q47 = the school average for the weekly teacher-reported hours students spend in class practicing writing; q52 = the school average for teacher years of teaching experience; q53 = the school average for teacher years of experience teaching writing; frl = percentage of students eligible for free or reduced-price lunch; cons = intercept; sigma_u = school-level random effect; sigma_e = student-level residual; rho = conditional ICC. Parameter estimates for the pair fixed effects were omitted from the table.
Source: Authors' analysis, based on data described in text.

Appendix M. Technical note on pooling of effect estimates

The two effect estimates, calculated separately from paired schools and from singleton schools, were pooled using the inverse-variance weighting method. This method is commonly used for a fixed-effect meta-analysis, and allows the pooling of the parameter estimates and the standard errors. The test statistic, *z*, can be calculated from these, and then used for the significance test. The following describes this pooling procedure in detail:

Suppose that a_1 is the parameter estimate from paired schools, and a_2 is that from singleton schools.

(1) Calculate the pooled impact estimate M using the inverse-variance weighting:

$M = (w_1 * a_1 + w_2 * a_2) / (w_1 + w_2)$, where $w_i = 1 / var(a_i)$

(2) Calculate the variance for the pooled impact estimate M:

$var(M) = 1 / (w_1 + w_2)$

The standard error of the estimate for the significance test of M, SE(M), is the square root of var(M).

(3) Calculate a *z*-value:

$z = M / SE(M)$

(4) Use the normal cumulative distribution to perform a significance test. Two-tailed tests were used for all the significance tests in this report.

References

ACT. (2005). *Crisis at the core: preparing all students for college and work*. Retrieved April 10, 2010, from http://www.act.org/research/policymakers/pdf/crisis_report.pdf

Applebee, A.N. (1986). Problems in process approaches: Toward a reconceptualization of process instruction. In A.R. Petrosky and D. Bartholomae (Eds.), *The teaching of writing: eighty-fifth yearbook of the National Society for the Study of Education, part II* (pp. 95–113). Chicago: University of Chicago Press.

Arter, J.A., Spandel, V., Culham, R., and Pollard, J. (1994, April). *The impact of training students to be self-assessors of writing*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED370975)

Bangert-Drowns, R.L., Hurley, M.M., and Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Review of Educational Research, 74*(1), 29–58.

Bartko, J.J. (1991). Measurement and reliability: Statistical thinking considerations. *Schizophrenia Bulletin, 17*(3), 483–489.

Berkner, L., and Choy, S. (2008). *Descriptive summary of 2003–04 beginning postsecondary students: three years later* (NCES 2008–174). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved July 5, 2010, from http://nces.ed.gov/pubs2008/2008174.pdf

Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–74.

Bloom, H.S. (Ed.). (2005). *Learning more from social experiments: Evolving analytic approaches.* New York: Russell Sage Foundation.

Bloom, H.S. (2007, July). *Sample design for group-randomized trials.* PowerPoint presentation at the Institute for Education Sciences, National Center for Education Research Summer Research Institute, Vanderbilt University, Nashville, TN.

Boscolo, P. (2008). Writing in primary school. In C. Bazerman (Ed.), *Handbook of research on writing: history, society, school, individual, text* (pp. 293–309). New York: Lawrence Erlbaum.

Casner-Lotto, J., and Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st*

century U.S. workforce. New York: Conference Board. Retrieved April 10, 2010, from http://www.p21.org/documents/FINAL_REPORT_PDF09-29-06.pdf

Casner-Lotto, J., Rosenblum, E., and Wright, M. (2009). *The ill-prepared U.S. workforce: exploring the challenges of employer-provided workforce readiness training*. New York: Conference Board. Retrieved April 10, 2010, from http://www.shrm.org/Research/SurveyFindings/Articles/Documents/BED-09Workforce_RR.pdf

Coe, M.T. (2000). *Direct writing assessment in action: correspondence of six-trait writing assessment scores and performance on an analog to the Washington Assessment of Student Learning writing test.* Portland, OR: Northwest Regional Educational Laboratory.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cole, N.S. (1997). *The ETS gender study: how females and males perform in educational settings.* Princeton, NJ: Educational Testing Service. (ERIC ED424337)

Common Core State Standards Initiative. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects.* Retrieved February 10, 2011, from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438–481.

Culham, R. (2003). *6+1 traits of writing: The complete guide, grades 3 and up*. New York: Scholastic Professional Books.

Diederich, P.B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Emig, J. (1971). *The composing processes of twelfth graders*. Urbana, IL: National Council of Teachers of English.

Executive Office of the President of the United States, Council of Economic Advisers. (2009). *Preparing the workers of today for the jobs of tomorrow*. Retrieved April 10, 2010, from http://www.whitehouse.gov/administration/eop/cea/Jobs-of-the-Future/

Fitzgerald, J., and Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35*(1), 39–50.

Flay, B.R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15*(5), 451–474.

Flower, L., and Hayes, J.R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387.

Gilmore Research Group. (2006). *Summary of proceedings from the 2006 state education forums.* Portland, OR: Northwest Regional Educational Laboratory.

Gilmore Research Group. (2009). *2007/2009 regional needs assessment summary of findings.* Portland, OR: Northwest Regional Educational Laboratory.

Graham, S., and Hebert, M. (2010). *Writing to read: evidence for how writing can improve reading. A report from Carnegie Corporation of New York.* Washington, DC: Alliance for Excellent Education.

Graham, S., and Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476.

Graham, S., and Perin, D. (2007b). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools. A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Harris, K.R., and Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation* (2nd ed.). Cambridge, MA: Brookline Books.

Hillocks, G., Jr. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: National Council of Teachers of English.

Hillocks, G., Jr. (1987). Synthesis of research on teaching writing. *Educational Leadership, 44*(8), 71–76, 78, 80–82.

Johnson, R.L., Penny, J., Gordon, B., Shumate, S.R., and Fisher, S.P. (2005). Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores? *Language Assessment Quarterly, 2*(2), 117–146.

Keys, C.W. (2000). Investigating the thinking processes of eight grade writers during the composition of a scientific laboratory report. *Journal of Research in Science Teaching, 37*(7), 676–690.

Kozlow, M., and Bellamy, P. (2004). *Experimental study on the impact of the 6+1 Trait® writing model on student achievement in writing*. Portland, OR: Northwest Regional Educational Laboratory. Retrieved July 27, 2010, from http://educationnorthwest.org/webfm_send/134

Marzano, R.J. (2003). *What works in schools: Translating research into action.* Alexandria, VA: Association for Supervision and Curriculum Development.

McGraw, K.O., and Wong, S.P. (1996). Forming inferences about some intraclass

correlation coefficients. *Psychological Methods, 1*(1), 30–46.

National Commission on Writing. (2003). *The neglected "R:" The need for a writing revolution*. Retrieved April 10, 2010, from http://www.collegeboard.com/prod_downloads/writingcom/neglectedr.pdf

National Commission on Writing. (2004). *Writing: A ticket to work…or a ticket out. A survey of business leaders*. Retrieved April 10, 2010, from http://www.collegeboard.com/prod_downloads/writingcom/writing-ticket-to-work.pdf

National Commission on Writing. (2005). *Writing: A powerful message from state government*. Retrieved April 10, 2010, from http://www.collegeboard.com/prod_downloads/writingcom/powerful-message-from-state.pdf

National Commission on Writing. (2006). *Writing and school reform*. Retrieved April 10, 2010, from http://www.collegeboard.com/prod_downloads/writingcom/writing-school-reform-natl-comm-writing.pdf

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist, 22*(2), 155–175.

Nowell, A., and Hedges, L.V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: an analysis of differences in mean, variance, and extreme scores. *Sex Roles, 39*(1/2), 21–43.

Oregon Department of Education. (2005). *Writing—the scoring process*. Retrieved July 29, 2010, from http://www.ode.state.or.us/teachlearn/subjects/elarts/writing/assessment/scoringprocess.pdf

Parsad, B., and Lewis, L. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000* (NCES 2004–010). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved October 25, 2010, from http://nces.ed.gov/pubs2004/2004010.pdf

Persky, H.R., Daane, M.C., and Jin, Y. (in collaboration with Davis, S., Jenkins, F., Liu, H., Qian, J., and Thind, S.). (2003). *The nation's report card: Writing 2002* (NCES 2003–529). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved July 27, 2010, from http://nces.ed.gov/nationsreportcard/pdf/main2002/2003529.pdf

Pritchard, R.J., and Honeycutt, R.L. (2006). The process approach to writing instruction: Examining its effectiveness. In C.A. Macarthur, S. Graham, and J. Fitzgerald (Eds.), *Handbook of writing research* (pp.275–290). New York: Guilford Press.

Purves, A.C. (Ed.). (1988). *Writing across languages and cultures: Issues in contrastive rhetoric.* Newbury Park, CA: Sage.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.

Salahu-Din, D., Persky, H., and Miller, J. (2008). *The nation's report card: writing 2007. National Assessment of Educational Progress at grades 8 and 12: national, state, and trial urban district results* (NCES 2008–468). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved July 27, 2010, from http://nces.ed.gov/nationsreportcard/pdf/main2007/2008468.pdf

Santa, C.M., Havens, L., Nelson, M., Danner, M., Scalf, L., and Scalf, J. (1988). *Content reading including study systems: Reading, writing and studying across the curriculum*. Dubuque, IA: Kendall Hunt. (ERIC Document Reproduction Service No. ED372363)

Scardamalia, M., Bereiter, C., and Fillion, B. (1981). *Writing for results: A sourcebook of consequential composing activities.* Toronto, Ontario, Canada: OISE Press.

Schochet, P.Z. (2005). *Statistical power for random assignment evaluations of education programs*. Princeton, NJ: Mathematica Policy Research.

Shanahan, T. (2004). Overcoming the dominance of communication: writing to think and to learn. In T.L. Jetton and J.A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 59–73). New York: Guilford Press.

Shrout, P.E., and Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

Sperling, M., and Freedman, S.W. (2001). Research on writing. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 370–389). Washington, DC: American Educational Research Association.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2008). *What Works Clearinghouse: Procedures and standards handbook* (Version 2.0). Retrieved July 8, 2010, from http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf

U.S. Department of Education, National Center for Education Statistics. (2003a). *The Nation's Report Card: state writing 2002. Snapshot report: Oregon grade 4, public school* (NCES 2003-532OR4). Retrieved February 11, 2011, from http://nces.ed.gov/nationsreportcard/pdf/stt2002/writing/2003532OR4.PDF

U.S. Department of Education, National Center for Education Statistics. (2003b). *The Nation's Report Card: state writing 2002. Snapshot report: Oregon grade 8, public*

school (NCES 2003-532OR8). Retrieved February 11, 2011, from http://nces.ed.gov/nationsreportcard/pdf/stt2002/writing/2003532OR8.pdf

U.S. Department of Labor, Secretary's Commission on Achieving Necessary Skills. (1993). *Teaching the SCANS competencies*. Retrieved July 27, 2010, from http://wdr.doleta.gov/SCANS/teaching/teaching.pdf


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA