The Challenge of Improving Children's Writing Ability: A Randomized Evaluation of *Writing Wings*

Michael Puma
Allison Tarkow
Anna Puma

Prepared for the Institute of Education Sciences, US Department of Education

December 2007

Acknowledgements

The authors would like to first thank all of the teachers and school administrators who graciously agreed to participate in this research study. And, of course, we are grateful to all of the students who cooperated with our data collection and shared their wonderful stories with us. Without them this project would not have been possible.

Second, we would like to thank two individuals who spent time on the road visiting schools and conducting the student data collection – Dr. David Connell and Terry Murray – as well as Dan Hoople who helped with the initial Year 1 data processing and analysis. And all of the teachers who served as readers for the student essays: Adelaide (Anne) Schoenhut, Janet Klingaman, Diane Rotella, Margaret Brady, Virginia Applebaum, and V. Patricia Currie.

We would also like to express our gratitude to the Institute of Education Sciences (IES) for their financial support, and to Stefanie Schmidt, our Project Officer, for her continued support and guidance.

Finally, we would like to thank our colleagues at the Success for All Foundation for help with site recruitment and the implementation of the *Writing Wings* curriculum at the selected study schools. In particular, we could not have done this study without the assistance of Nancy Madden, Michele Karanzalis, Denise Douce, and all of their field trainers who worked with the study teachers.

Executive Summary

Study Objectives

The ability to write is a fundamental part of literacy, and a skill that all children need to acquire for success in school and throughout life. As noted by the National Writing Project (2003), "Writing is the gateway to success in school, helping students learn to read, to solve problems, and to understand concepts in every part of the curriculum. It is also the currency of the new workplace and global economy where it often has to be produced instantly and effectively." As important as writing is for success both in school and in later life, recent data from the National Assessment of Educational Progress (NAEP) show that fewer than one in three 4th-graders, one in three 8th-graders, and one in four 12th-graders scored at or above the proficient level. These data also indicate persistent and substantial achievement gaps between white and African American students, and between white and Hispanic students.

To address this need, this study evaluated the effectiveness of a structured writing program for 3rd, 4th, and 5th graders, a critical period in the acquisition of good writing skills. The instructional program that was studied, called *Writing Wings*, was developed by the Success For All Foundation (SFAF) to meet the learning needs of all students, but especially the needs of disadvantaged students. The curriculum is research-based and seeks to enhance teachers' skills and enable them to succeed at teaching their students to write through a combination of clear instructional goals, teacher modeling, and a cooperative writing process.

The study sought to answer the following research questions about the effect of *Writing Wings*:

- The overall confirmatory question is, "What is the impact of *Writing Wings* on the writing ability of 3rd, 4th, and 5th grade elementary students?"
- In addition two more exploratory questions were addressed, "What is the impact of *Writing Wings* on the writing attitudes and practices of 3rd, 4th, and 5th grade elementary students?" and "Do the impacts vary by the characteristics of the students and their teachers?"

Study Design

This cluster randomized control trial (RCT), involved two separate annual cohorts of high-poverty elementary schools – 17 schools in Year 1 (2005-06) and a new sample of 22 schools in Year 2 (2006-07). At each school in Year 1, two 3rd grade and two 4th grade classrooms were randomly selected and assigned to either: (1) a **treatment group** in which *Writing Wings* was implemented, or (2) to a **control group** in which students received whatever instruction was currently in place in their particular schools (i.e., the intervention is compared to the "business as usual" model of instruction). The same procedure was implemented in Year 2, but with the addition of a 5th grade cohort. (All control group teachers were provided with *Writing Wings* training after they completed their involvement in the impact study.) To the best of our knowledge, this is the only experimental study that has been done of a elementary school writing instructional

program since research on the Cognitive Strategy Instruction in Writing in 1991 (Englert, et al., 1991).

Comparable data were collected from all students in both the randomly assigned treatment and control group classrooms. Initial or baseline data were collected, to the extent possible, during the first few weeks of school in Fall 2005 and Fall 2006 (for Cohorts 1 and 2, respectively) with follow-up data collected in Spring 2006 (Cohort 1) and Spring 2007 (Cohort 2). All data were collected by study staff who conducted visits to each of the selected schools.

The primary data collection activity involved the administration of a structured writing test to each student. Children were randomly given one of two writing "prompts" – a **narrative** assignment involving a personal experience, or an **informative** assignment requiring the students to share knowledge and information about a favorite game/sport, movie, or book. Children were allowed, on average, 40 minutes to complete the writing task. In addition, a questionnaire was used to collect information about student's demographic characteristics, knowledge of writing mechanics, and reported writing "confidence." A teacher questionnaire was also administered to collect information about teacher and classroom characteristics. In addition, teachers were asked to rate each student's writing ability in both Fall and Spring. Finally, implementation fidelity data were collected in each treatment classroom, and observations were conducted in the control classrooms to check on any possible spillover of the *Writing Wings* instruction.

The combined randomly assigned sample consisted of about 3,000 students associated with 152 classrooms teachers in 39 schools located in 21 different states. The final impact analysis sample consists of the same 39 schools, but the student sample was reduced to 2,405 students, a very high overall 80 percent response rate.[1] Accordingly, the impact analysis was run both with and without weights intended to adjust the post-test sample back to the initial randomly assigned sample of students.

Findings

There is evidence of some impact on teacher's instructional practices with regard to writing. Teachers in the treatment condition reportedly taught writing more often and for longer sessions, had students execute multiple aspects of the writing process more frequently, and felt more confident teaching writing than did teachers in the control group.

However, despite these indications of impacts on intermediate outcomes, there were no overall statistically significant impacts on student's writing ability, and a similar lack of statistically significant impacts on teacher ratings of their student's writing ability. However, student responses to the survey indicated statistically significant differences on several individual questionnaire items, and the multivariate analysis confirms that *Writing Wings* appears to have had a relatively large positive impact on the frequency of student-reported in-school writing. Looking at the possibility of impacts on selected subgroups of students, these more exploratory analyses indicate that there may be an

[1] Student mobility was the primary reason for the pre- to post-test attrition, but other events also affected the final sample including one school that was re-organized and the mobilization of military families at a school associated with a military base.

impact on student's written organization skills for 5th grade students, and at least a suggestion of a positive impact on the frequency of in-school writing for those students who had relatively low initial writing scores.

Conclusions

These results are disappointing but may not be surprising in light of several conditions may have contributed to the observed lack of impacts on student outcomes:

1. Teachers were reportedly at a fairly high level of writing instruction at the start of the study, and the observed effects on teaching practices may have been too small to have a large enough subsequent impact on student outcomes to be detected with this ample size.
2. The initial training of teachers was delayed for Cohort 1, and implementation data indicate that the time required for teachers to reach an acceptable level of implementation was slower than originally expected. Consequently, the expectation that a single year of experience with the new curriculum would be sufficient to reach the required level of teacher proficiency may have been incorrect.
3. Further, the level of teacher training and support may have been inadequate to create the necessary change in the quantity and quality of writing instruction that students received.
4. Finally, students' initial writing performance was generally low; students scored, on average, a "2" or "3" on the 5-point essay scale at the time of the baseline assessment. Thus, teachers in the current study were instructing many students with substantial need in terms of writing skills.

Writing Wings is a complicated program that requires a significant amount of effort on the part of classroom teachers to reach a desirable level of instructional practice. Not only is there a lot for teachers to absorb and to figure out how to best integrate the instruction into their existing classroom routines, but early feedback from teachers indicates that they often struggled with finding the added time needed to bring this level of writing instruction into a schedule that is already stretched.

As a consequence, SFAF should take these results as an opportunity to re-evaluate their instructional program, and the associated professional development model, to find a way to create the level of change in instructional practice needed to improve student's writing ability. In addition, in light of the relatively low level of student writing ability that was observed in the Fall, SFAF should consider exposing students to the program for consecutive years (e.g., 3rd and 4th grades), and possibly integrating the instruction across subject areas, to get the desired improvement on student achievement.

This is not to say that impacts would have been different if the program and the training had been implemented differently, but at least these suggested changes would remove some of the possible reasons why the estimated effects were so weak thereby increasing our confidence in any observed impacts on student outcomes.

Table of Contents

List of Exhibits

Chapter 1: Background and Study Overview

Writing well is not just an option for young people—it is a necessity. Along with reading comprehension, writing skill is a predictor of academic success and a basic requirement for participation in civic life and in the global economy.
(Graham & Perin, 2006)[2]

Writing Skills are Critical For Student Success

The ability to write is a fundamental part of literacy, and a skill that all children need to acquire for success in school (Strech, 1994).[3] Good writers tend to do well in reading (Strickland, 1991;[4] Teale & Sulzby, 1989;[5] National Reading Panel, 2000;[6] National Research Council, 1997, 1998[7]), and there are important "spillover" effects into other subjects such as mathematics, social studies, and science. That is, becoming a good writer provides students with the tools they need to learn a variety of subjects (Keys, 2000;[8] Shanahan, 2004;[9] Sperling & Freedman, 2001[10]).

Learning to be a good writer goes beyond an understanding of the mechanics of grammar, syntax, and vocabulary, as it requires the ability to organize one's thoughts and present a convincing argument, as well as the creativity to develop stories and poems. These are skills that can make a student successful in all instructional content areas and throughout his or her life. As noted by the National Writing Project (2003),[11] "In this new century, writing is a skill of increasing importance. Writing is the gateway to success in school, helping students learn to read, to solve problems, and to understand concepts in every part of the curriculum. It is also the currency of the new workplace and global economy where it often has to be produced instantly and effectively."

---

[2] Graham, S. & D. Perin (2006). ***Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High School.*** New York, NY: The Carnegie Corporation of New York.

[3] Strech, L.L. (1994). "The Implementation of Writing Workshop: A Review of the Literature." *ERIC* Document ED 380 797.

[4] Strickland, D.S. (1991). "Emerging Literacy: How Young Children Learn To Read." In B. Persky And L.H. Golubchick (Eds.), ***Early Childhood Education*** (2nd ed.) (337-344). Lanham, MD: University Press of America.

[5] Teale, W.H., & Sulby, E. (1989). "Emergent literacy: New perspectives." In D.S. Strickland and L. M. Morrow (Eds.), ***Emerging Literacy: Young Children Learn To Read And Write.*** Newark, DE: International Reading Association.

[6] National Reading Panel (2000). ***Teaching Children To Read: An Evidence-Based Assessment Of The Scientific Research Literature On Reading And Its Implications For Reading Instruction.*** Rockville, MD: National Institute of Child Health and Human Development.

[7] National Research Council (1997). *Improving Schooling For Language-Minority Children*. Diane August & Kenji Hakuta, Editors. Washington, DC. National Research Council (1998). *Preventing Reading Difficulties in Young Children.* Catherine Snow, M. Susan Burns, & Peg Griffin, Editors. Washington, DC, 1998.

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

[9] Shanahan,T. (2004). "Overcoming the dominance of communication: Writing to think and to learn." In T. L. Jetton & J. A. Dole (Eds.). ***Adolescent Literacy Research And Practice*** (pp. 59–73). NewYork: Guilford.

[10] Sperling, M., & Freedman, S.W. (2001). "Review of writing research." In V. Richardson (Ed.), ***Handbook Of Research On Teaching*** (4th ed., pp. 370–389).Washington, DC:American Educational Research Association.

[11] National Writing Project (2003). "National Writing Project Statement on the 2002 NAEP Writing Report." http://www.writingproject.org/pub/nwpr/news/2003/2002naep.html.

For some, writing so closely resembles thinking and learning that it is seen as a concrete manifestation of general cognitive development (Emig, 1977).[12] According to this view, to write is to learn (Bangert-Drowns, et al., 2004).[13] As noted by the National Writing Project (2003), "Writing is the process by which we learn how to convey our ideas, to use our powers of observation, and to persuade others about our viewpoints…..If writing occurred in every classroom every day, student achievement across content areas would reach new heights for all."

But There is a Problem of Poor Achievement in Writing

As important as writing is for success both in school and in later life, there are persistent and substantial achievement gaps between white and African American students, and between white and Hispanic students. And the achievement gaps appear to worsen as children progress through school.

In particular, recent data from the National Assessment of Educational Progress (NAEP)[14] show improvements in the overall writing performance of America's 4th graders from 1998 to 2002, but the gap between white and African American, between white and Hispanic children (the minority children also tend to be concentrated in high-poverty schools), and between students eligible for free/reduced-price lunch and those who are ineligible, remained substantial, especially for older students (NAEP, 2003). In fact, fewer than one in three 4th-graders, one in three 8th-graders, and one in four 12th-graders scored at or above the proficient level. In addition, students in urban fringe/large town schools posted higher writing scores than students in both central city and rural schools, and girls exhibited higher scores than boys. The worsening of performance for the 12th graders is not surprising as evidence suggests that early literacy differences tend to grow over the years and are difficult to remediate once they are well established (Juel, 1988).[15]

The extant literature has identified many components of good writing instruction, but unfortunately disadvantaged students ─ those most in need of high-quality instruction ─ are far less likely than other students to have teachers who are well qualified to teach writing or who spend adequate classroom time on the subject (College Board, 2003).[16] In fact, most 4th graders spend less than three hours per week writing, a statistic that prompted the National Commission on Writing to recommend that schools incorporate writing into the instruction of every student in the elementary grades (College Board, 2003).

Clearly, if the achievement gap is to be significantly closed, disadvantaged and minority children must have an effective writing program as part of their formal schooling. According to evidence from NAEP, students scored higher on the writing assessment if

---

[12] Emig, J. (1977). "Writing as a Mode of Learning." ***College Composition and Communication***, 28, 122-128.

[13] Bangert-Drowns, R.L., Hurley, M.M., & 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.

[14] National Assessment of Educational Progress (2003). ***NAEP Writing, 2002***. Washington, DC: National Center for Education Statistics.

[15] Juel, C. (1988). "Learning to read and write: A longitudinal study of 54 children from first through fourth grades." ***Journal of Educational Psychology****, 80,* 437-447.

[16] College Board (2003). ***The Neglected 'R': The Need For A Writing Revolution.*** Report of the National Commission on Writing in America's Schools and Colleges. New York, New York.

they reported having discussed their writing with their teachers, and if their teachers frequently asked them to plan their writing in advance and to write more than one draft. In addition, engaging in prewriting activities and writing in a log or journal were also correlated with higher writing achievement.

Study Goals

To address this discrepancy in student performance, this study evaluated the effectiveness of a structured writing program for 3rd, 4th, and 5th graders, a critical period in the acquisition of good writing skills. The instructional program, called *Writing Wings*, was developed by the Success For All Foundation (SFAF) to meet the learning needs of disadvantaged students. The program is "ready to go" and can be readily replicated across the country, serving as a model for other schools serving disadvantaged students and helping to narrow the noted achievement gap. SFAF's programs are currently being implemented in thousands of schools spread across all parts of the United States, Puerto Rico, Guam, and the Virgin Islands, and versions of their programs are also used in other countries, including England, Israel, Canada, Mexico and Australia.

The study sought to answer the following research questions about the effect of *Writing Wings*:

- The overall confirmatory question is, "What is the impact of *Writing Wings* on the writing ability of 3rd, 4th, and 5th grade elementary students?"
- In addition two more exploratory questions were addressed, "What is the impact of *Writing Wings* on the writing attitudes and practices of 3rd, 4th, and 5th grade elementary students?" and "Do the impacts vary by the characteristics of the students and their teachers?"

The Writing Wings Intervention

Writing Instruction in Elementary School. The teaching of writing has undergone a substantial revolution since the 1970's (see Calkins, 1983;[17] Graves, 1983;[18] Dyson & Freedman, 1991;[19] and Freedman, et al., 1987[20]). Where writing instruction once focused substantially on the mechanics of grammar and punctuation, the current focus is on helping students gain insight into the writer's craft (Harris & Graham, 1996).[21] In particular, "writing process" models emerged in the 1970's in which children are taught to plan, draft, revise, edit, and publish compositions in various genres (De La Paz & Graham, 1997;[22] Graham, 1997[23]). The theory behind these methods is that through the

[17] Calkins, L. (1983). ***Lessons From a Child***. Portsmouth, NH: Heinemann.
[18] Graves, D. H. (1983). ***Writing: Teachers and Children at Work***. Portsmouth, NH: Heinemann.
[19] Dyson, A., & Freedman, S. (1991). "Writing." In J. Flood, J. Jensen, D. Lapp, & J. Squire (Eds.), ***Handbook Of Research On The Teaching The English Language Arts.*** (pp. 787-802). New York: Macmillan.
[20] Freedman, S., Dyson, A., Flower, L., & Chafe, W. (1987). ***Research In Writing: Past, Present, And Future***. Berkeley, CA: Center for the Study of Writing.
[21] Harris, K.R., & Graham, S. (1996). ***Making The Writing Process Work: Strategies For Composition And Self-Regulation***. Cambridge, MA: Brookline.
[22] De La Paz, S., & Graham, S. (1997). "Effects of dictation and advanced planning instruction on the composing of students with writing and learning problems." ***Journal of Educational Psychology***, *89*, 203-222.
[23] Graham, S. (1997). "Executive control in the revising of students with learning and writing difficulties." ***Journal of Educational Psychology***, *89*, 223-234.

process of revision, editing, and rewriting, students will gain insight into their own writing skills (Bereiter & Scardamalia, 1987;[24] Case, et al., 1994[25]). Cooperative writing teams, or peer response groups, are typically used in these models, and it is hypothesized that reviewing others' writing and hearing others discuss their thinking about writing helps young authors begin to comprehend the underlying structure of composition and to progressively improve their own writing products. In fact, studies of writing process models in comparison to other strategies for teaching composition have found effects favoring the writing process strategies both on the quality of student compositions (rated on various rubrics) as well as language mechanics, such as usage and punctuation (Hillocks, 1984;[26] Harris & Graham, 1996[27]).

While there has been a great deal of descriptive research about writing instruction, there is remarkably little experimental research on practical approaches to teaching this critical skill. A key exception is research on Cognitive Strategy Instruction in Writing (CSIW; Englert, et al., 1991).[28] This method emphasizes daily writing, motivational procedures, and explicit instruction on writing structures (such as comparison/contrast), "think alouds," to verbally model thinking about composition, and construction and revision of text. Research on CSIW has found significant positive effects on students' writing skills (Englert. et al., 1991).

Writing Wings. The *Writing Wings* instructional writing program is research-based drawing heavily on the research and experience of the CSIW project, but specifically seeks to target the unique needs of underserved populations and high-poverty schools. The aim of *Writing Wings* is to enhance teachers' skills and enable them to succeed at teaching their students to write through a combination of clear instructional goals, teacher modeling, and a cooperative writing process.

Writing Wings consists of twelve units of detailed daily lesson plans that support instruction in descriptive, informative, persuasive, and narrative writing. Each unit includes writing prompts and writing challenges, with instruction spanning over ten days. Within units, lessons highlight the writing process of planning, drafting, sharing and responding, revising, and editing. The lessons include language mechanics and are rich with cooperative learning opportunities. The writing challenges that are embedded in the program assess students on their individual writing. The challenges help teachers and students measure and celebrate progress in writing, and help identify weaknesses at both the classroom and student level. Each writing challenge includes a writing prompt and scoring rubric. Teachers may choose to use the included *Writing Wings* rubric, or grade their students' writing based on the rubric used in their state, if one exists.

---

[24] Bereiter, C., & Scardamalia, M. (1987). ***The Psychology of Written Communication*****.** Hillsdale, NJ: Erlbaum.

[25] Case, L., Mamlin, N. Harris, K., & Graham, S. (1994). "Self regulated strategy development: A theoretical and practical perspective." In T. Scruggs and M. Mastropieri (Eds.), ***Advances In Learning And Behavioral Disabilities*****.** Greenwich, CT: JAI Press.

[26] Hillocks, G. (1984). "What works in teaching composition: A meta-analysis of experimental treatment studies." ***American Journal of Education***, *93*, 133-170.

[27] Harris, K.R., & Graham, S. (1996). Making The Writing Process Work: Strategies For Composition And Self-Regulation. Cambridge, MA: Brookline.

[28] Englert, C.S., Raphael, T.E., Anderson, L.M., Anthony, H.M., & Stevens, D.D. (1991). "Making strategies and self-talk visible: Writing instruction in regular and special education classrooms." ***American Educational Research Journal***, *28*, 337-372.

For the second study Cohort (see Chapter 2), a recently modified version of *Writing Wings* was introduced that most notably added videos to accompany most lessons. This medium is not only intended to engage students in learning, but also models critical skills and concepts. The videos use puppets and animated characters to model various parts of the writing process, to teach language mechanics and grammar lessons, and to demonstrate cooperative learning skills.

Teacher Training and Support: SFAF staff were involved in the initial recruitment of potential study sites (see Chapter 2) and in the implementation of the *Writing Wings* intervention. They had no role in random assignment, impact study data collection, or the analysis of the data.

With regard to the implementation of the *Writing Wings* intervention, for both years of the study SFAF trainers were scheduled to contact and visit research sites several times a year. These points of contact consisted of an initial 1-day on-site training for the teachers who were randomly assigned to the treatment group, followed by three support visits and four telephone support meetings spread over the school year.

The initial training took place at the beginning of the school year after baseline testing was completed (see Chapter 2). Teachers were provided a comprehensive introduction to *Writing Wings*, and teachers were able to familiarize themselves with the curriculum by incorporating cooperative learning concepts, the cycle of effective instruction, ways of monitoring student progress, and target setting.

Three support visits were scheduled throughout the year, augmented with a total of four follow-up support calls. The typical schedule of training contacts, after the initial training session, was as follows: an initial follow-up support call shortly after the training, a visit and a support call in September-December, a second visit and support call in January-March, and a final support call in April-May. Visits and calls were oriented to helping teachers progress in their implementation of the curriculum. As part of the visits, SFAF trainers observed and provided feedback, and addressed any questions or concerns teachers had regarding the curriculum or its implementation. Feedback sessions also consisted of planning reachable targets for the next support visit.

The Pace of Implementation: As part of the three support visits, SFAF staff conducted observations of each treatment and control group classroom (the latter was done to look for any possible spillover effects). In the treatment classes the raters assessed the current state of program implementation on a 4-point scale. The results of these "fidelity of implementation" ratings show that teachers required some time to reach an acceptable level of classroom implementation:

- Approximately two months after initial training, about half (52%) of the treatment classrooms scored a 3 or 4 on the implementation scale (indicating "routine" or "refined" implementation).
- Approximately five months after initial training, 69 percent of the treatment classrooms scored a 3 or 4 on the implementation scale.

- At the end of the school year – just before the collection of post-test data – nearly three-quarters (72%) of the treatment classrooms scored a 3 or 4 on the implementation scale.

The implications of this observed pace of implementation for the program impact estimates is discussed in Chapter 5.

Chapter 2: Study Design and Methods

Study Design

As noted in Chapter 1, the purpose of this study is to determine whether *Writing Wings* has an **impact** on the writing ability of elementary school children. By impact we mean a difference between the outcomes observed for children who receive this type of instruction and what *would have been observed for these same students had they **not** participated in Writing Wings*. Unlike other education studies that often seek to examine relationships between participant outcomes and one or more individual or program characteristics, the proposed study seeks to assess the extent to which the intervention – *Writing Wings* – **caused** any observed student outcomes.

Given this goal of measuring program impacts, how do we determine what outcomes would have been observed if the children had not participated in *Writing Wings*? That is, how do we observe students having the same characteristics in two places at the same time **—** in and not in *Writing Wings* **—** and compare them? In many studies, researchers have addressed this problem by comparing program participants to a "participant-like" group of children. However, even the best attempts at constructing such a comparable group of non-participants suffer from what evaluators call "selection bias." That is, children or teachers who are "selected" to participate in the writing program may be different from those who do not on important factors that may lead to different outcomes independently of the effect of the instructional program itself. For example, teachers who might decide on their own to implement the program (or who effectively "lobby" their principal) may be more motivated and they may have a class of higher ability children than those who opt not to do so. Moreover, these factors are both typically unobserved and likely to be related to the outcomes of interest in their own right. That is, the motivated teachers may do a host of things that may affect their student's achievement levels beyond providing them with the new instructional program.

To avoid this problem of selection bias, this study was designed as a cluster randomized control trial, involving two separate annual cohorts of high-poverty elementary schools – 17 schools in Year 1 (2005-06) and a new sample of 22 schools in Year 2 (2006-07).[29] At each school in Year 1, two 3rd grade and two 4th grade classrooms were randomly selected and assigned to either: (1) a **treatment group** in which *Writing Wings* was implemented, or (2) to a **control group** in which students received whatever instruction was currently in place in their particular schools (i.e., the intervention is compared to the "business as usual" model of instruction). The same procedure was implemented in Year 2, but with the addition of a 5th grade cohort. (All control group teachers were provided with *Writing Wings* training after they completed their involvement in the impact study.)

[29] The original study design planned to have a total of 20 schools with four teachers per school, and to follow the students for two years. That is, the students initially randomly assigned in Fall 2005 were to be followed through Spring 2007; students assigned to the respective treatment and control group classrooms were to remain intact and continue in their same assignment condition for the next school year (i.e., treatment children would receive *Writing Wings* for two years). This plan, however, proved infeasible as it was not possible to maintain intact classes of students across years. Because of this, and the shortfall in the expected Year 1 sample, we decided to add a new set of schools, teachers, and students in Year 2. The two cohort samples were pooled for analysis.

Under this randomized design, a simple comparison of outcomes for the two groups yields an unbiased estimate of the impact of *Writing Wings* on children's writing ability. The advantage of this research design is that if random assignment is properly implemented with a sufficient sample size, program participants should not differ in any systematic or unmeasured way from non-participants except through their access to the new instructional program. More precisely, there will be differences between individuals in the two groups, but the expected or average value of these differences is zero except through the influence of *Writing Wings* (i.e., selection bias is removed by random assignment).

This certainty of attribution to the right causal factor can never be achieved if schools and staff make their own choices about what type of instruction gets used; too much else about the students and teachers will differ. That is, unequivocal answers on "what works?" can only be obtained where unequivocal comparisons give a result attributable to one deliberately-varied factor with, in statistical terms, all else held equal.

Although researchers have suggested a large number of non-experimental methods for achieving the same purpose such as multivariate regression, selection correction methods (Heckman & Hotz, 1989),[30] and propensity score methods (Rosenbaum & Rubin, 1983),[31] a long line of literature, including recent analyses by Bloom, et al. (2002),[32] Agodini & Dynarski (2001),[33] and Wilde & Hollister (2002),[34] suggests that none of these methods is as reliable as random assignment.

The estimated power for the study design was calculated to be approximately 0.18 standard deviation units, i.e., the study had an estimated 80 percent chance of detecting a statistically significant treatment-control group difference on student writing achievement measures if the true differences were at least 0.18 effect size units.

Sample Selection and Recruitment: Sites were recruited from within the thousands of schools around the country that were currently using the Success for All (SFA) instructional program. SFA schools use cooperative learning in a structured literacy program to support reading achievement, but the Foundation had not yet offered schools a complete writing program.

For Cohort 1, the recruitment process was begun immediately upon notification of the award in June 2005 for implementation in September 2005, but moved slowly as schools

[30] Heckman, J.J. & Hotz, V.J., 1988. "*Choosing Among Alternative Nonexperimental Methods For Estimating The Impact Of Social Programs: The Case Of Manpower Training*," University of Chicago, Economics Research Center 88-12.

[31] Rosenbaum P.R., Rubin D.B. (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." ***Biometrika*** 70, 41-55

[32] Bloom, H., et al., (2002). *Can Non-Experimental Comparison Group Methods Match the Findings From a random Assignment Evaluation of mandatory Work-to-Welfare Programs?* New York, NY: MDRC.

[33] Agodini, R. & Dynarski, M. (2001). "Are Experiments the Only Option? A look at Dropout Prevention programs." ***The Review of Economics and Statistics***, 86(1), 190-194.

[34] Wilde, E.T. & Hollister, R. (2002). "How Close Is Close Enough? Testing Nonexperimental Estimates of Impact against Experimental Estimates of Impact with Education Test Scores as Outcomes." Department of Economics, Swarthmore College.

were closed for the summer;[35] for Cohort 2, site recruitment began in January 2006, for implementation in August 2006.

To recruit schools for the study, SFAF staff (including headquarters research staff, field personnel, Area Managers, and trainers) contacted schools both directly and as part of their participation in the annual SFA Experienced Site Conferences. Additional recruitment actions included creating and posting an ad on SFAF's intranet for current schools, "fax blasts," and announcements in newsletters routinely distributed to schools. A total of approximately 175 schools indicated an initial interest in participating in one of the two cohorts.

Each interested school was contacted by a member of the research staff and provided with extensive information about the requirements for participation, including the requirement that teachers be randomly assigned to treatments, and a description of the materials and support that would be provided. Schools indicating continued interest were followed until their questions were answered and they were enrolled in the study, or until they declined to participate. For many schools, the requirement that only half of the teachers participate was not acceptable as they wished all of their teachers to be working with the same program. Other schools had two teachers interested, but would not accept random assignment to treatment. In general, the recruitment discussion took 4 or 5 contacts to complete, and about a fourth of the schools indicating interest ultimately joined the study.

The study design was carefully explained to potential study sites, and all sites were required to have a Research Agreement signed in advance by both the school principal and the District Superintendent. The agreement specified the random assignment process and the planned data collection; teachers assigned to the control group were to be provided with the *Writing Wings* training and materials at the end of the study.

Once a school agreed to participate, parental consent forms were delivered to the school for distribution to parents during the first week of school. Baseline data collection was then scheduled for as soon as possible after a reasonable period of time to collect the signed consent forms.[36] (Copies of the Research Agreement and parent consent form are provided in Appendix A.)

Random Assignment: Only after baseline data were collected were schools informed of which teachers were randomly assigned either to the treatment or control group – this was done to reduce the likelihood of schools organizing classrooms in a way that might affect the equivalence across the two study groups. For the most part, random assignment was conducted and implemented as planned but in three instances school administrators undermined the initial assignment by sending the control teacher to the *Writing Wings*

[35] The research grant was awarded mid-June 2005 with the expectation that the evaluation would still start by September 2005 (the originally proposed date). Although sites were initially recruited as part of the grant proposal process, this late date required the study team to essentially start the recruitment process over again. By mid-July 2005 SFAF staff had assembled a list of 22 "highly interested" study sites. An intensive effort was then implemented to recruit the schools with the eventual inclusion of 17 schools in Year 1.

[36] An example of how unexpected events can affect a study such as this was the disaster of Hurricane Katrina which hit while staff were out collecting baseline data for Year 1. This led to the loss of a school from the study and unplanned disruptions in the timing of baseline data collection and initial teacher training at two other schools in Year 1.

training. In each case, because this was uncovered after the collection of baseline data, the grade pair – the paired treatment and control classes – were dropped from the study.

Data Collection

Comparable data were collected from all students in both the randomly assigned treatment and control group classrooms in all of the selected study schools. Initial or baseline data were collected, to the extent possible, during the first few weeks of school in Fall 2005 and Fall 2006 (for Cohorts 1 and 2, respectively) prior to the start of any *Writing Wings* teacher training, with follow-up data collected in Spring 2006 (Cohort 1) and Spring 2007 (Cohort 2). All data were collected by study staff who conducted visits to each of the selected schools.[37]

The primary data collection activity involved the administration of a structured writing test to each student. Children were given one of two writing "prompts" (see Exhibit 1) – a **narrative** assignment involving a personal experience, or an **informative** assignment requiring the students to share knowledge and information about a favorite game/sport, movie, or book. The prompts were randomized within each class of students, and the same prompts were used for both the fall and spring administrations – because the prompts were randomized at each data collection, we did not try to ensure that any individual student received either the same or a different writing prompt. Children were allowed, on average, 40 minutes to complete the writing task.

Each essay was "blind coded" with respect to the students treatment assignment by two different experienced elementary school teachers who were specially trained in the scoring rubric used for this study. Each rater independently scored each essay on a 1-5 scale (1/2 points were allowed) on four different dimensions:

- **Content and Audience Awareness –** the ability to engage the reader in the selected topic exhibiting a grasp of the details. Is it interesting?
- **Organization –** the use of proper sequencing (a beginning and end) and flow. Does it make sense?
- **Word Choice –** the use of interesting and varied words to convey meaning and to capture the reader's interest.
- **Conventions –** proper us of grammar, paragraphing, punctuation, and spelling

The scoring rubric provided a definition for three "anchor" points for each dimension: a score of 1.0 (the lowest), a score of 3.0 (the mid-point) and a score of 5.0 (the highest score). Teachers were trained using examples of actual student writing corresponding to each of these three scores for each of the four dimensions.

In the event that two ratings differed by more than 1.0 points, an independent third rater scored the essay (this was usually the master trainer). Each student's final essay rating was the average of the two (or in some cases three) ratings on each of the four dimensions. In addition, an overall total score was also computed as the average score across the four sub-scales.

[37] In a few instances, school staff were asked to complete student data collection for absent students. This represented, however, a small number of students.

Exhibit 1: Student Writing Prompts

Prompt 1: Narrative

Carefully read the prompt in the box below. Read it again to be sure you understand it before you start. You have 40 minutes to write your story in the space below the box. You can use the section at the back to get your ideas together. Be sure to review your work and check spelling.

> Select a place you have been (such as the beach, a park, a movie theater, the grocery store) and describe your experience.

Prompt 2: Informative

Carefully read the prompt in the box below. Read it again to be sure you understand it before you start. You have 40 minutes to write your story in the space below the box. You can use the section at the back to get your ideas together. Be sure to review your work and check spelling.

> Choose **one** of the following and write a story that will tell the reader "all about" the topic you select:
>
> - Your favorite game or sport.
> - A favorite book.
> - A favorite TV show or movie.

The assessment appears to function as would be expected. First, as shown below for the Year 1 Cohort, 4th graders scored higher than 3rd graders (a statistically significant difference) and Spring scores were higher than initial fall scores, both indications of expected growth over time:

	3rd Grade		*4th Grade*	
Component	***Fall 2005***	***Spring 2006***	***Fall 2005***	***Spring 2006***
Average Score	1.76	2.29	2.12	2.53
Word Choice	1.72	2.27	2.08	2.50
Content	1.88	2.43	2.27	2.68
Organization	1.73	2.30	2.10	2.55
Conventions	1.71	2.15	2.03	2.41

Further, on average girls scored higher than boys, and white students scored higher than African American and Hispanic students.

In addition to the writing assignment, a brief 25 item questionnaire was used to collect information about student's demographic characteristics, knowledge of writing mechanics, and reported writing "confidence" (see Appendix B). Field staff read the questionnaire to the students so that even students with low reading skills could fill it out correctly. A teacher questionnaire (consisting of a total of 34 items and also provided in Appendix B) also collected information about teacher and classroom characteristics. In addition, teachers were also asked to rate each student's writing ability in both fall and spring.

The student and teacher questionnaires were developed by the research team for this study based on the existing literature on good writing instruction. The student questionnaire items used to create outcome measures were focused on three different domains: (1) attitudes about writing and student's motivation to write well (predictors of student engagement and achievement); (2) use of the various components of the "writing process" that form the basis for *Writing Wings* (see Chapter 1) for in-school writing; and, (3) student's understanding of the criteria used to grade their school writing, also a key component of good writing instruction. The teacher questionnaire focused on four key domains: (1) teacher background characteristics (education, certification, years of experience); (2) characteristics of writing instruction including frequency of instruction per week, the length of time used for writing instruction, grading and assessment, and the components of the writing process that are used and their frequency; (3) integration of writing across subject areas; and, (4) teacher's confidence to teach writing. The questionnaire items were adapted from, or modeled after, items used in other similar surveys including the student questionnaires from the National Assessment of Educational Progress (NAEP).

Finally, implementation fidelity data were collected by SFAF staff in each treatment classroom at the same time that the trainers were on site for the scheduled training/support visits using the *Writing Wings* Progress Logs, which were updated

during support visits with the combined efforts of observations and teacher/trainer conferences. Writing samples from both treatment and control classes were collected. Trainers collected samples of high, medium, and low student performance on similar activities or time frames from each grade level participating in the study. Trainers assigned a rating to each teacher assigned to the *Writing Wings* treatment for each support visit. The ratings were on a scale ranging from one to four. A rating of one equaled little or no implementation, while a rating of four equaled refined implementation. Control teachers were also observed during support visits to ensure against treatment overspill.

Impact Analysis Methods

Outcome Measures. The following measures were used as outcomes for all of the student impact analyses (all measured in the spring of the two study years):

- *Writing Ability – Essay Test:*
    - **Total Average Essay Rating** − Average of the four separate essay scores described below − Content and Audience Awareness, Word Choice, Organization, and Conventions. Each dimension was assessed on 1 to 5 scale (5 = exemplary).
    - **Essay Rating: Content and Audience Awareness** − Measures the student's ability to engage the reader in the selected topic exhibiting a grasp of the details. Is the paper interesting?
    - **Essay Rating: Word Choice** – Measures the student's use of interesting and varied words to convey meaning and to capture the reader's interest.
    - **Essay Rating: Organization** – Measures the student's use of proper sequencing (a beginning and end) and flow. Does the writing make sense?
    - **Essay Rating: Conventions** − Measures the student's proper use of grammar, paragraphing, punctuation, and spelling
- *Teacher Rating of Student Writing Ability* − Teacher rating of student's relative writing ability assessed on a 1 to 5 scale (5 = well above average).
- *Student Writing Attitudes Scale (SWA)* − A composite score based on 11 individual items from the student survey related to their writing attitudes (Questions 2A-E and 3A-F from the questionnaire). Total scores ranged from 11 to 44 where each item was rated on "how true" they were from 1 to 4 (1 = Really True).
- *Student In-School Writing Scale (SSW)* − A composite score based on 7 individual items from the student survey related to their in-school writing (Question 4A-G). Total scores ranged from 7 to 21 where items were rated on the frequency of occurrence from 1 to 3 (1 = Most of the time).
- *Student Writing Grades Scale (SWG)* − A composite score based on 3 individual items from the student survey related to how student perceptions of what it takes to achieve good writing grades (Question 5A-C). Total scores ranged from 3 to 9 where items were rated on importance from 1 to 3 (1 = Very Important).

Missing baseline values (used as covariates in the impact analyses described below) were imputed with the mean for other students in same school, grade, and treatment status.

Overall Impact Analysis. The overall impact questions for this study are, "What is the impact of *Writing Wings* on the writing ability of 3rd, 4th, and 5th grade elementary students?" and "What is the impact of *Writing Wings* on the writing attitudes and practices of 3rd, 4th, and 5th grade elementary students?" As noted above, all analyses used to answer these questions were conducted using the pooled Cohort 1 and Cohort 2 data sets.

Hierarchical linear modeling (HLM)[38] was used to answer these questions for two main reasons: (1) the sample was selected in a way that reflects the nesting of students within classrooms and classrooms within schools, and HLM allows for the correct calculation of variances (and associated statistical significance) with such a clustered sample; and (2) multi-level analysis allows examination of classroom context effects (i.e., treatment) on student performance over and above prior learning experiences at an individual level.

The actual model construction was done separately for each of the outcomes listed above using a 3-level hierarchical linear model: Level 1 was estimated at the student level, Level 2 was estimated at the classroom level, and Level 3 was estimated at the school level. As such: Level 1 parameters explain why students vary on a given outcome; Level 2 parameters explain how classroom characteristics, especially the assignment to treatment, affects student outcomes; and, the final Level 3 parameters estimate the variability between schools.

In each model, the fall student level variable for that outcome was entered into the model as a control. For instance, when the Total Essay Rating score was entered as the outcome, the Fall baseline measure of this same variable was entered at Level 1 as a control for students' previous performance. Additionally, because factors other than the writing program could account for student writing performance, important student characteristics (i.e., gender, race) were also entered as covariates at Level 1. All control variables were centered on the grand mean and thus generalize to all students across schools (see Exhibit 2 for a list of the covariates used in each of the models).

The general model for ***Level 1*** is given below:

$$\text{SpringY}_{ijk} = \pi_{0jk} + \pi_{1jk}(\boldsymbol{FALL}) + \pi_{2jk}(\boldsymbol{MALE}) + \pi_{3jk}(\boldsymbol{BLACK}) + \pi_{4jk}(\boldsymbol{HISPANIC}) + \pi_{5jk}(\boldsymbol{OTHER}) + e_{ijk}$$

(*Note: ***GRAND MEAN CENTERED,*** **GROUP MEAN CENTERED,** NOT CENTERED*)

In each model, the treatment status of the classroom was entered at Level 2, modeled on the intercept. Similarly, cohort status and grade level of the classroom was entered at Level 2, modeled on the intercept. Level 2 predictors were not centered to facilitate interpretability of the model estimates. These coefficients indicate, for example, the added benefit of being in a treatment, cohort 2, 4th and 5th grade classroom for spring writing performance.

[38] See, for example, Bryk, A.S. & S.W. Raudenbush (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.

The general model for ***Level 2*** is given below:

$\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{TREATMT}) + \beta_{02k}(\text{COHORT2}) + \beta_{03k}(\text{GRADE4}) + \beta_{04k}\text{GRADE5} + r_{0jk}$

$\pi_{1jk} = \beta_{10k}$

$\pi_{2jk} = \beta_{20k}$

$\pi_{3jk} = \beta_{30k}$

$\pi_{4jk} = \beta_{40k}$

$\pi_{5jk} = \beta_{50k}$

Lastly, the school level variance is estimated at level 3 of the model (no predictors were included at the school level and homogeneity among classrooms was assumed). The general model for ***Level 3*** is: $\beta_{00k} = \gamma_{000} + u_{00k}$

Thus, the ***combined or overall model*** is as follows:

$\text{SpringY}_{ijk} = \gamma_{000} + \beta_{01k}(\text{TREATMT}) + \beta_{02k}(\text{COHORT2}) + \beta_{03k}(\text{GRADE4}) + \beta_{04k}(\text{GRADE5}) + \beta_{10k}(\textbf{\textit{FALL}}) + \beta_{20k}(\textbf{\textit{MALE}}) + \beta_{30k}(\textbf{\textit{BLACK}}) + \beta_{40k}(\textbf{\textit{HISPANIC}}) + \beta_{50k}(\textbf{\textit{OTHER}}) + e_{ijk} + r_{0jk} + u_{00k}$

(*Note: ***GRAND MEAN CENTERED,*** **GROUP MEAN CENTERED,** NOT CENTERED*)

All analyses were conducted using the pooled Cohort 1 and Cohort 2 data sets.

Exhibit 2: Covariates Used in the Impact Analyses

Student Level Predictors	
Male	Child gender is Male=1; Female=0
Black	If child is African American, Black=1, else Black=0
Hispanic	If child is Hispanic, Not Black Hispanic=1 else Hispanic=0
Other	All other children are coded Other=1, else Other=0
Baseline Measure of the Outcome	The Fall measure corresponding to the particular outcome variable being examined.
Classroom-level Predictors	
Treatment	Treatment Status=1, Control Status=0
Cohort2	If child was in Cohort 2, then Cohort2 =1, else Cohort2=0 (i.e., in Cohort 1)
Grade4	If child is in 4th grade, Grade4=1, else Grade4=0
Grade5	If child is in 5th grade, Grade5=1, else Grade5=0 (3rd grade is excluded category)

Subgroup Impacts. The final exploratory research question posed for this study is, "Do the impacts vary by the characteristics of the students and their teachers?"

To address this question, additional analyses were conducted by adding interaction terms to the model above where a particular subgroup indicator (e.g., MALE) was interacted with the treatment group indicator variable (MALE*Treatment). The parameter estimate for these interaction terms indicate whether there is a statistically significant impact on the particular outcome (e.g., Total Essay Score) for the group used to define the interaction term – in this example, is there a statistically significant impact on student's Total Essay Score for males?

The subgroups examined in these analyses were as follows:

- *Students:*
    - **Gender** – impacts on boys (girls are the excluded category);
    - **Grade level** – impacts on 4th and 5th grade students (3rd grade is the excluded category).
    - **"Low" achieving students** – impacts on students who scored either a 1 or a 2 (out of 5) on their fall baseline essay (students not categorized as low are the excluded category).
- *Teachers:*
    - **Teacher Experience** – teachers with less than 5 years of experience (5 or more years of experience is the excluded category).
    - **Teacher Reported Baseline Instructional "Confidence"** – teachers who reported "low" levels of confidence in teaching writing (those not indicating low confidence are the excluded category).

The statistical models were developed separately for each of the outcomes listed above using a 3-level hierarchical linear model as was done for the overall main impacts. As with the main impact analyses, all models were estimated using the pooled Cohort 1 and Cohort 2 data sets.

Weights. Because of a concern about the loss of about 20 percent of the randomly assigned sample of students (see Chapter 3), all of the above analyses were replicated with the use of non-response weights, i.e., the post-test sample was weighted based on the response rate for each of the study schools. (The non-response weights were included at Level 3.) The HLM results were run with and without weights to adjust for post-test attrition. The coefficient parameters were similar in magnitude and direction between the weighted and un-weighted models, although the estimates were more reliable using the weighted version. Consequently, the results reported in Chapter 4 use the results of the weighted models.

Chapter 3: Study Sample

Sample Overview

As discussed above, for Cohort 1, the baseline student sample was split between the 3rd and 4th grades (47% and 53%, respectively) with 51 percent of students randomly assigned to the treatment (*Writing Wings*) group and 49 percent to the control group (random assignment was done by class and not by individual students). For Cohort 2, the baseline student sample was split among the 3rd, 4th and 5th grades (40%, 41% and 19%, respectively); the smaller 5th grade cohort was a result of a greater difficulty getting schools to agree to add a new instructional program at this grade level for this study. As with Cohort 1, the sample was about evenly split by treatment group (51% assigned to the treatment group and 49% to the control group).

The initial Cohort 1 sample consisted of 17 schools spread across 11 states, and a total baseline sample 1,179 3rd and 4th grade students and 68 classroom teachers; the initial Cohort 2 sample consisted of 22 schools spread across 11 states (a total of 21 different states across the two cohorts), and a total baseline sample 1,827 3rd, 4th and 5th grade students and 84 classroom teachers. The total initial baseline sample (combining both cohorts), therefore, consisted of 3,006 students in 39 different schools.

This sample was quite diverse:

- The 21 states were located nationally and included the northeast, mid-atlantic, south, southeast, midwest, southwest, west, and northwest regions.
- The schools were located in major US cities, suburban communities, and in sparsely populated rural areas. One school was on an American Indian Reservation, and one was located on a US Army Military base.
- The average school size was 437 students (range: 200-700);
- Free and reduced-price meal eligibility ranged from 30 to 100 percent, with a mean of 73 percent.
- The average school was 28 percent white (range, 0-95%), 39 percent African American (range, 0-100%), 19 percent Hispanic (range, 0-94%), 3 percent Native American (range, 0-100%), and 11 percent Asian American (range, 0-96%).

The final impact analysis sample consists of the same 39 schools, but the student sample was reduced to 2,405 students, a very high overall 80 percent response rate. Student mobility was the primary reason for the pre- to post-test attrition but other events also affected the final sample including one school that was re-organized and the mobilization of military families at a school associated with a military base. As discussed in Chapter 2, the impact analysis was run both with and without weights intended to adjust the post-test sample back to the initial baseline.

Data collection consisted of both a student survey and a structured writing assignment. Survey data were also collected from students' classroom teachers, and teachers were also asked to rate the students' writing ability. The following describes the baseline

characteristics of the final impact sample, with a particular focus on any observed statistical differences between the treatment and control groups.

Impact Sample -- Baseline Characteristics

Student Demographics: As shown in Exhibit 3, 49 percent of the impact sample is male and 51 percent are female, with no statistically significant differences between the treatment and control groups.

About 43 percent of the combined Cohort 1-Cohort 2 impact sample is African American, about 27 percent is white, about 17 percent is Hispanic, with "Other Races" (primarily Asian/Pacific Islander and Native American) making up the remaining 14 percent of the sample. As shown, the treatment group contains a higher percentage of Hispanic students than the control group ($p<0.01$). As noted in Chapter 2, such initial differences are controlled for in the impact analysis by the inclusion of baseline covariates.

Teacher Ability Ratings: Teachers were asked to rate each student's writing ability using the following prompt, "**Compared to other students, this student's writing ability is….**" where students were rated on a 5-point scale: 1=well below average, 2=below average, 3=average, 4=above average, and 5=well above average. As also shown in Exhibit 3, students were ranked relatively low by their teachers at the start of the school year with no statistically significant differences between the treatment and control groups.

Student Scales: Student survey data were used to create three separate scales used as outcome variables in the impact analyses:

- *Writing Attitudes Scale (SWA)*, measuring the extent to which they enjoy writing;
- *School Writing Scale (SSW)*, measuring students' writing practices within school such as using charts to organize thoughts or writing multiple drafts; and,
- *Writing Grades Scale (SWG)*, which included various questions about what was required to get a good grade in a writing assignment (e.g., writing an interesting paper and making few spelling errors).

As shown in Exhibit 4, students reported moderately positive attitudes towards writing, somewhat frequent writing, and somewhat positive notions of obtaining good writing grades. Small but statistically significant differences were found on two of the scales with students in the treatment group reporting writing more frequently, and holding more positive ideas of what determines writing grades compared to students in the control group. Again, these small differences are controlled for in the impact analysis.

Exhibit 3: Student Baseline Characteristics by Treatment Group, Combined Cohort 1 and 2 Impact Sample, Fall 2005 and Fall 2006

Characteristic	TOTAL	Treatment	Control	Difference
Male	49.1%	49.8%	48.3%	1.5%
Race/Ethnicity:				
White	26.6%	25.7%	27.5%	-1.8%
Black	43.0%	42.9%	43.0%	-0.1%
Hispanic	16.5%	17.9%	15.1%	**2.8%****
Other Races	13.8%	13.4%	14.2%	-0.8%
Teacher Rating of Writing Ability	2.64	2.61	2.67	-0.06

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 4: Student Baseline Writing Scales by Treatment Group, Combined Cohort 1 and 2 Impact Sample, Fall 2005 and Fall 2006

Component	TOTAL		Treatment		Control		Difference
	M	SD	*M*	SD	*M*	SD	
Writing Attitudes Scale (SWA; higher score indicates more positive attitudes)	23.75	4.62	23.62	4.62	23.89	4.62	-0.26
In School Writing Scale (SSW; lower score indicates more frequent use of the writing process)	12.57	2.45	12.41	2.40	12.74	2.50	**-0.33****
Writing Grades Scale (SWG; lower score indicates more positive perceptions of what constitutes good writing)	4.48	1.18	4.40	1.18	4.55	1.17	**-0.15****

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

In addition to the scales, the two experimental groups were also compared on the individual survey items. For the pooled impact sample, the baseline survey indicates that study children:

- Reportedly wrote relatively frequently – nearly 40 percent report writing every day or "most days," and about 80 percent responded "true" or "really true" to the statement "I like to write."
- Had positive perceptions of the importance of writing – nearly 80 percent responded "true" or "really true" to the statements, "Writing helps me to think more clearly," and to the statement, "People who write well do better in school."
- Saw themselves as good writers (despite their low teacher assessments and, as discussed below, their low performance on the essay test) – 78 percent respond "true" or "really true" to the statement "I am a good writer," and 72 percent respond "true" or "really true" to the statement say "People like what I write."

When they wrote in school, many of the students in both cohorts were reportedly exposed to the "writing process" at the time of the baseline survey, including the use of webs or charts "most of the time" to organize their thoughts (32%), writing multiple drafts (35%), sharing their writing with a partner (35%), revising their writing based on feedback (37%), checking their work for spelling and punctuation (57%), and sharing their work with others (27%).

Across the 22 total questionnaire items, there were, however, a few statistically significant differences at baseline:

- Students in the control group were slightly more likely than students in the treatment group to feel they would not write at all if it weren't for school (36% of the control group said "really true" or "true" to this statement compared to 32% in the treatment group; $p<.05$).
- Students reported the use of several strategies when writing in class, however, students in the treatment group reported exposure to the "writing process" somewhat more often than the control group for a few items. For instance, students in the treatment group were reportedly more likely than the control group to write multiple drafts (88% of the treatment group said "really true" or "true" to this statement compared to 84% in the control group; $p<.05$), revise their writing (85% of the treatment group said "really true" or "true" to this statement compared to 80% in the control group; $p<.01$), and practice punctuation or grammar using worksheets (85% of the treatment group said "really true" or "true" to this statement compared to 79% in the control group; $p<.001$).
- Finally, students reported what was required to get a good grade in a writing assignment including writing an interesting paper and making few spelling errors. Students appeared to understand what was important for achieving high grades rating these items as important. However, students in the treatment group were reportedly more likely than the control group to report that writing an interesting paper (95.2% vs. 94.5%; $p<.05$) and making few spelling errors (81% vs. 78%; $p<.05$) were more important for getting a good grade than students in the control group.

The reader is cautioned, however, to not place too much emphasis on these results because of the relatively large number of statistical tests that were run on the individual survey items. Under such conditions of "multiple comparisons," there is a modest probability that a finding of a statistically significant difference will emerge by random chance — an event known in the statistical field as a "false discovery." False discoveries are most common when *p*-values are very close to the prescribed cut-off level for statistical significance, in this case $p < .05$, and when many comparisons are made, thereby giving random chance extra opportunities to produce an apparently (but not actually) statistically significant finding.

Student Essays: The final component of the data collection, and certainly the most important as it forms the primary outcome measure for the study, was the administration of a standardized essay writing test to the students. As shown in Exhibit 5, student writing ability was generally low both in terms of the overall rating and for each of the individual dimensions. As one would expect, students in the 4th grade scored higher than students in the 3rd grade, and students in the 5th grade scored higher than the other students. Most importantly, there were no observed statistically significant treatment-control differences as of the time of the baseline assessment.

The baseline essay scores also exhibited high variability across the 39 schools in the combined impact sample. One school fell at the very bottom of the distribution with the following averages across the students in the study (combined treatment and control at baseline): Total Average Score – 1.41, Word Choice – 1.39, Content – 1.53, Organization – 1.36, and Convention – 1.35. The school at the top of the distribution on all dimensions had the following average ratings: Total Average Score – 2.65, Word Choice – 2.51, Content – 2.76, Organization – 2.82, and Convention – 2.51. Furthermore, the "low end" school had about 83 percent of the students receiving a 1 or a 2 on the total average score, compared to about 27 percent for the "high end" school.

Baseline Teacher Characteristics

A survey was also administered to students' teachers in the Fall of 2005 and 2006 to obtain information regarding: their educational background and teaching experience; the extent to which they focus on teaching writing to their students; grading criteria used for writing; instructional methods used to teach writing; and, their level of comfort to teach writing. Across the two cohorts for the impact sample, we see the following picture at the time of the baseline survey:

- **Teacher qualifications:** Over half of the teachers had only a BA/BS degree, and over 1/3rd had a graduate degree. On average, teachers had about 11 years of teaching experience.
- **Writing Emphasis:**
  - Nearly 7 out of 10 teachers reportedly taught writing at least 3 days/week, and about 40 percent did so every day.
  - When they taught writing, about 75 percent of teachers reportedly spent 30-60 minutes on writing (11% reported spending more than one hour).

Exhibit 5: Student Baseline Essay Ratings by Treatment Group, Combined Cohort 1 and 2 Impact Sample, Fall 2005 and Fall 2006

Component	TOTAL		Treatment		Control		Difference
	M	**SD**	***M***	**SD**	***M***	**SD**	
Total Score:	**2.01**	**.84**	**2.02**	**.83**	**2.01**	**.84**	**0.01**
3rd Grade	1.76	.72	1.72	.70	1.80	.75	-0.08
4th Grade	2.12	.84	2.17	.84	2.07	.83	0.10
5th Grade	2.56	.88	2.55	.87	2.57	.89	-0.02
Word Choice Score:	**1.97**	**.83**	**1.97**	**.82**	**1.97**	**.85**	**0.00**
3rd Grade	1.72	.72	1.68	.69	1.76	.75	-0.08
4th Grade	2.08	.82	2.13	.81	2.03	.83	0.10
5th Grade	2.50	.91	2.48	.88	2.51	.95	-0.03
Content and Audience Awareness Score:	**2.15**	**.94**	**2.14**	**.95**	**2.15**	**.94**	**-0.01**
3rd Grade	1.88	.84	1.83	.82	1.93	.86	-0.10
4th Grade	2.27	.96	2.31	.96	2.22	.95	0.09
5th Grade	2.67	.94	2.66	.95	2.68	.93	-0.02
Organization Score:	**2.00**	**.91**	**2.00**	**.90**	**1.99**	**.91**	**0.01**
3rd Grade	1.73	.78	1.70	.74	1.77	.82	-0.07
4th Grade	2.10	.91	2.15	.92	2.05	.89	0.10
5th Grade	2.59	.98	2.58	.97	2.61	.98	-0.03
Conventions Score:	**1.94**	**.88**	**1.95**	**.88**	**1.94**	**.88**	**0.01**
3rd Grade	1.71	.77	1.67	.75	1.75	.79	-0.08
4th Grade	2.03	.90	2.08	.90	1.98	.90	0.10
5th Grade	2.48	.89	2.49	.86	2.47	.91	0.02

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

- **Instructional Practice:**
  - In terms of grading, about 50 percent of teachers reported giving separate grades for content and mechanics. Across both cohorts, there was a reportedly strong grading focus on the completion of the assignment, organization, and the use of complete sentences; somewhat less importance was reported by the teachers in the areas of the use of interesting language, punctuation, spelling, and neatness.
  - In terms of instructional methods, teachers reported the following for either "every" or "most" lessons: 85 percent spent time teaching prewriting or drafting skills; 70 percent taught editing skills; 2/3rd taught revising skills; and about half had students share and publish their work.
  - The key "writing process" steps were reportedly "always" used by many teachers: 80 percent had students write a first draft, and 60 percent had them use graphic organizers and spend time revising their draft. Fewer teachers reportedly "always" used other components of the writing process: proofreading and making final edits (53%), revising drafts based on feedback (36%), publishing their work (34%), sharing drafts with other students (27%), and working with a partner on developing their written piece (18%).
  - Nearly every teacher claimed to incorporate writing instruction into other subjects, particularly social studies, science, and math; and, about half said they incorporated writing instruction into the arts.
  - Nearly 7 out of ten teachers reportedly had their students keep a writing portfolio.
- **Teacher Confidence:** When asked, "How comfortable do you feel teaching writing?" about 1/5th of the teachers said they were "somewhat uncertain," 35 percent said "comfortable," and 45 percent said "pretty confident" or "very confident."

Across all of the teacher survey items, there were no statistically significant baseline differences between the treatment and control group.

A final question on the survey asked an open-ended question of the teachers, "What do you find the most difficult part of writing instruction?" The following is a summary of the areas noted by all teachers who took the time to respond across both cohorts (number of responses):

- **Insufficient time to teach writing:** This was probably the most commonly expressed concern both anecdotally and on the teacher surveys (38).
- **Organization:** Getting students to develop and organize their ideas (32); getting students understand and write to the prompt staying on topic (10).
- **Content:** Getting students to use more descriptive language (5), getting students to write creatively and with imagination– telling a story – entertaining the reader (3), getting students to include good supporting details (6).
- **Mechanics:** Getting students to use complete sentences (7), working on spelling, grammar and punctuation (9).

- **Editing and Revision:** Getting students to proofread, edit, and revise (28).
- **Student Motivation:** Motivating students to work hard and put effort into their work (8).
- **Dealing with Varying Ability levels:** Dealing with students with different levels of ability (17).
- **Other Concerns:** Finding materials to use in the classroom (2), getting a published piece of writing (1), conferencing with other students (2), grading student writing (4), integrating writing into other instructional areas (1).

The next chapter presents the results of the impact analyses, and the final Chapter 5 discusses the results.

Chapter 4: Impact Results

This chapter presents the overall impact results followed by the estimated impacts for selected subgroups of students. Chapter 5 discusses the implications of these results.

Main Impact Findings

Teachers

By the time of the Spring follow-up survey, there appears to have been some impact on teacher's instructional practices with regard to writing. As shown in Appendix C, teachers in the treatment condition reportedly taught writing more often and for longer sessions, had students execute multiple aspects of the writing process more frequently (e.g., sharing writing, publishing, using worksheets, working with partners, using graphic organizers, and sharing and revising drafts), and felt more confident teaching writing than did teachers in the control group (all statistically significant at $p \leq 0.05$).

Impact on Student Writing Ability

Exhibits 6 and 7 provide the estimated impacts on the student essays measures – Exhibit 6 provides simple mean differences, while Exhibit 7 provides the results of the HLM analysis. As shown, despite the reported changes in teaching practices *Writing Wings* was found to have no statistically significant impact on children's writing ability either overall, or by the four writing components. And, as indicated by the effect sizes shown in Exhibit 6, the estimated treatment-control differences are quite small.

In reviewing these results, readers should keep in mind that the criteria used for determining the statistical significance of any individual impact estimate means the reported treatment-control group difference may have arisen by pure chance when no true effect occurs only 5 out of 100 times, i.e., when an impact is detected, we can be very confident that we have observed a real non-zero impact. Where instead, as in these results, we fail to find a statistically significant difference (i.e., we fail to reject the null hypothesis of no impact on a particular outcome) we do not have conclusive evidence that the program "doesn't work." Rather, statistically insignificant impacts mean that the effect is indeterminate – *Writing Wings* may or may not have had a non-zero impact on student's writing ability, and we cannot with this study sample make a confident conclusion either way. The one thing that will be known with confidence is that a large true impact has not occurred.

Exhibit 6: Mean Student Essay Ratings by Treatment Group, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007

Essay Component and Grade Level	TOTAL		Treatment		Control		Difference (Effect Size)
	M	**SD**	***M***	**SD**	***M***	**SD**	
Total Score:	**2.45**	**.81**	**2.46**	**.80**	**2.44**	**.82**	**.021 (0.03)**
3rd Grade	2.29	.76	2.28	.74	2.30	.78	-.017
4th Grade	2.53	.83	2.56	.82	2.50	.84	.064
5th Grade	2.78	.77	2.77	.80	2.79	.76	-.021
Word Choice Score:	**2.43**	**.80**	**2.45**	**.80**	**2.41**	**.81**	**.045 (0.06)**
3rd Grade	2.27	.75	2.28	.74	2.27	.76	.009
4th Grade	2.50	.82	2.55	.81	2.46	.83	.089
5th Grade	2.74	.80	2.73	.83	2.74	.78	-.007
Content and Audience Awareness Score:	**2.60**	**.89**	**2.60**	**.89**	**2.59**	**.90**	**.009 (0.01)**
3rd Grade	2.43	.85	2.42	.83	2.45	.87	-.031
4th Grade	2.68	.91	2.70	.91	2.65	.91	.054
5th Grade	2.92	.85	2.90	.88	2.93	.83	-.029
Organization Score:	**2.46**	**.87**	**2.47**	**.86**	**2.46**	**.89**	**.012 (0.01)**
3rd Grade	2.30	.83	2.29	.80	2.31	.87	-.021
4th Grade	2.55	.89	2.58	.89	2.52	.90	.059
5th Grade	2.76	.84	2.73	.85	2.79	.83	-.063
Conventions Score:	**2.33**	**.86**	**2.34**	**.85**	**2.32**	**.87**	**.021 (0.02)**
3rd Grade	2.15	.81	2.14	.79	2.16	.83	-.026
4th Grade	2.41	.88	2.44	.86	2.37	.90	.065
5th Grade	2.70	.81	2.70	.82	2.69	.82	.016

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 7: Impact of Treatment on Students' Writing Ability, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Total Score	Word Choice	Content Score	Organization Score	Conventions Score
Intercept, $\gamma 000$	2.460**	2.441**	2.601**	2.472**	2.331**
TREATMT, γ_{010}	0.005	0.040	-0.001	-0.012	0.006
COHORT2, γ_{020}	-0.044	-0.040	-0.008	0.001	-0.119
GRADE4, γ_{030}	0.070	0.096*	0.094	0.088	0.098*
GRADE5, γ_{040}	0.164**	0.184**	0.257**	0.194**	0.177*
FALL, $\gamma 100$	0.432**	0.376**	0.323**	0.359**	0.469**
MALE, $\gamma 200$	-0.162**	-0.165**	-0.198**	-0.184**	-0.162**
BLACK, $\gamma 300$	-0.183**	-0.212**	-0.204**	-0.180**	-0.156**
HISPANIC, $\gamma 400$	-0.177**	-0.202**	-0.195**	-0.198**	-0.125*
OTHER, $\gamma 500$	-0.009	-0.001	-0.040	-0.032	0.002
u_{0j}, Variance	0.033**	0.028**	0.050**	0.040**	0.026**
u_{00j}, Variance	0.030**	0.033**	0.032**	0.039**	0.027**
Intercept, $\gamma 00$, Reliability	0.528**	0.469	0.550	0.514	0.452
Intercept, $\gamma 000$, Reliability	0.623	0.661	0.555	0.632	0.614

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Impacts on Other Outcomes

As shown in Exhibits 8 and 9, there were also no statistically significant differences found for the teacher ratings of student's writing ability, a finding that confirms the lack of a significant difference on the student essay test.

However, student responses to the follow-up survey indicated statistically significant differences on several individual questionnaire items. Compared to students in control classrooms, students in the treatment group had more positive attitudes and confidence in their writing (e.g., I like writing, people like what I write), and reportedly engaged more frequently in various components of the writing process (e.g., use webs, write multiple drafts, share writing with a partner, revise writing, share completed work, practice grammar).

These mean differences on individual survey items are matched by noted statistically significant impacts on both the in-school writing scale and the writing grades scale, although only the former holds up with the HLM analysis (Exhibit 9). That is, *Writing Wings* appears to have increased the frequency of student-reported in-school writing, and the magnitude of this effect is quite large (effect size = 0.37 standard deviations).

Exhibit 8: Mean Student Teacher Ratings and Writing Scales by Treatment Group, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007

Component	TOTAL		Treatment		Control		Difference (Effect Size)
	M	SD	*M*	SD	*M*	SD	
Teacher rating of student writing ability	2.82	1.00	2.86	1.00	2.79	1.00	.070 (0.07)
Writing Attitudes Scale (higher score indicates more positive attitudes)	24.37	4.30	24.23	4.23	24.51	4.36	-0.279 (0.065)
In School Writing Scale (lower score indicates more frequent use of the writing process)	11.56	2.40	11.13	2.37	12.01	2.36	-0.879** (0.366)
Writing Grades Scale (lower score indicates more positive perceptions of what constitutes good writing)	4.47	1.19	4.42	1.13	4.53	1.25	-0.109* (0.092)

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 9: Impact of Treatment on Students' Writing Rating and Attitudes, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Teacher Rating	Writing Attitudes Scale	In School Writing Scale	Writing Grades Scale
Intercept, $\gamma 000$	2.815**	24.350**	11.488**	4.449**
TREATMT, γ_{010}	0.111	-0.130	-0.706**	-0.082
COHORT2, γ_{020}	0.034	-0.289	0.259	0.034
GRADE4, γ_{030}	0.146*	0.234	-0.144	-0.053
GRADE5, γ_{040}	0.157	0.741**	0.427	0.016
FALL, $\gamma 100$	0.638**	0.236**	0.222**	0.196**
MALE, $\gamma 200$	-0.159**	1.596**	0.427**	0.312**
BLACK, $\gamma 300$	-0.169**	-0.748**	0.076	-0.145
HISPANIC, $\gamma 400$	-0.173**	0.001	0.225	-0.009
OTHER, $\gamma 500$	-0.050	0.510	0.578**	0.197*
u_{0j}, Variance	0.097**	0.451**	0.614**	0.049**
u_{00j}, Variance	0.006	0.503**	0.130*	0.001
Intercept, $\gamma 00$, Reliability	0.702	0.299	0.653	0.353
Intercept, $\gamma 000$, Reliability	0.131	0.536	0.335	0.024

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Subgroup Impact Findings

Exhibits 10A through 10F and 11A through 11F provide the results of the analyses of subgroup impacts on the direct assessments of students' writing ability (Exhibits 10A-F), and on teacher ratings of students' ability and on the student scales of writing attitudes and behaviors (Exhibits 11A-F). (In these tables, FALLLow is the indicator of whether the student had a low Baseline essay score, TEXP is the subgroup of less than 5 years of teacher experience, and TCONF is the subgroup of low teacher-reported instructional confidence.)

With regard to the direct assessments of student's writing ability, as shown in Exhibits 10A-F there was a single statistically significant impact found for 5th grade students on the Organization score for their essay writing ($p \leq 0.05$). With respect to the other outcomes (Exhibit 11A-F), none of the estimated subgroup impacts met the chosen level of statistical significance ($p \leq 0.05$) but there are a couple of suggested effects ($p \leq 0.10$). There is an indication that *Writing Wings* has a positive impact on the frequency of in-school writing for those students who had relatively low initial writing scores, but a negative impact for boys on their perceptions of what it takes to get a good writing grade.

Exhibit 10A: Impact of Treatment on Students' Writing Ability for Gender, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Total Score	Word Choice	Content Score	Organization Score	Conventions Score
Intercept, γ000	2.430**	2.377**	2.539**	2.420**	2.319**
TREATMT, γ010	0.006	0.040	-0.001	-0.012	0.006
COHORT2, γ020	-0.045	-0.041	-0.008	0.001	-0.119
GRADE4, γ030	0.070	0.097*	0.095	0.088	0.098*
GRADE5, γ040	0.164**	0.184**	0.257**	0.194**	0.177*
FALL, γ100	0.433**	0.376**	0.323**	0.360**	0.470**
MALE, γ200	-0.193**	-0.206**	-0.219**	-0.209**	-0.191**
TRT*MALE, γ210	0.063	0.081	0.042	0.051	0.060
BLACK, γ300	-0.183**	-0.211**	-0.204**	-0.179**	-0.155**
HISPANIC, γ400	-0.177**	-0.203**	-0.195**	-0.199**	-0.126*
OTHER, γ500	-0.008	0.000	-0.039	-0.031	0.003
$u0_j$, Variance	0.033**	0.027**	0.049**	0.040**	0.026**
$u00_j$, Variance	0.030**	0.033**	0.032**	0.039**	0.027**
Intercept, γ00, reliability	0.524	0.465	0.549	0.512	0.447
Intercept, γ000, reliability	0.625	0.664	0.556	0.634	0.616

+ Statistical trend at the 90 percent confidence level;

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 10B: Impact of Treatment on Students' Writing Ability for 4th Grade, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Total Score	Word Choice	Content Score	Organization Score	Conventions Score
Intercept, γ000	2.442**	2.397**	2.551**	2.435**	2.334**
TREATMT, γ010	-0.017	0.003	-0.023	-0.041	-0.022
COHORT2, γ020	-0.046	-0.043	-0.009	-0.000	-0.122+
GRADE4, γ030	0.044	0.054	0.069	0.055	0.066
GRADE5, γ040	0.164**	0.185**	0.257**	0.194**	0.178*
TRT*GRADE4, γ050	0.053	0.086	0.051	0.068	0.064
FALL, γ100	0.432**	0.375**	0.323**	0.359**	0.469**
MALE, γ200	-0.162**	-0.165**	-0.198**	-0.184**	-0.162**
BLACK, γ300	-0.184**	-0.212**	-0.205**	-0.180**	-0.156**
HISPANIC, γ400	-0.178**	-0.204**	-0.196**	-0.200**	-0.126*
OTHER, γ500	-0.010	-0.002	-0.041	-0.033	0.001
$u0_j$, Variance	0.033**	0.027**	0.049**	0.040**	0.026**
$u00_j$, Variance	0.030**	0.034**	0.032**	0.039**	0.027**
Intercept, γ00, reliability	0.524	0.460	0.548	0.510	0.448
Intercept, γ000, reliability	0.628	0.670	0.559	0.638	0.618

+ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 10C: Impact of Treatment on Students' Writing Ability for 5th Grade, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Total Score	Word Choice	Content Score	Organization Score	Conventions Score
Intercept, γ000	2.424**	2.372**	2.533**	2.411**	2.316**
TREATMT, γ010	0.018	0.052	0.011	0.006	0.014
COHORT2, γ020	-0.044	-0.040	-0.008	0.001	-0.119+
GRADE4, γ030	0.070	0.096*	0.095	0.088	0.098*
GRADE5, γ040	0.224**	0.242**	0.318**	0.280**	0.217*
TRT*GRADE5, γ050	-0.118	-0.113	-0.121	**-0.168***	-0.078
FALL, γ100	0.432**	0.376**	0.323**	0.359**	0.469**
MALE, γ200	-0.162**	-0.166**	-0.199**	-0.184**	-0.162**
BLACK, γ300	-0.183**	0.212**	-0.204**	-0.180**	-0.156**
HISPANIC, γ400	-0.177**	-0.203**	-0.195**	-0.199**	-0.125*
OTHER, γ500	-0.008	-0.000	-0.039	-0.031	0.003
$u0_j$, Variance	0.033**	0.027**	0.049**	0.039**	0.026**
$u00_j$, Variance	0.030**	0.034**	0.032**	0.039**	0.027**
Intercept, γ00, reliability	0.524	0.464	0.548	0.508	0.450
Intercept, γ000, reliability	0.626	0.665	0.557	0.638	0.615

+ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 10D: Impact of Treatment on Students' Writing Ability for Low Fall Essay Scores, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Total Score	Word Choice	Content Score	Organization Score	Conventions Score
Intercept, γ000	2.430**	2.378	2.542**	2.419**	2.320**
TREATMT, γ010	0.006	0.040	-0.003	-0.012	0.006
COHORT2, γ020	-0.044	-0.040	-0.008	0.002	-0.119+
GRADE4, γ030	0.071	0.095*	0.090	0.088	0.098*
GRADE5, γ040	0.161**	0.183**	0.250**	0.196**	0.177*
FALL, γ100	0.484**	0.335**	0.271**	0.318**	0.466**
MALE, γ200	-0.161**	-0.165**	-0.197**	-0.184**	0.162**
BLACK, γ300	-0.181**	-0.212**	-0.207**	-0.182**	-0.156**
HISPANIC, γ400	-0.174**	-0.203**	-0.197**	-0.201**	-0.125*
OTHER, γ500	-0.011	0.005	-0.036	-0.029	0.003
FallLOW, γ600	0.097	-0.100	-0.122*	-0.122+	-0.011
TRT*FallLOW, γ610	0.011	0.029	-0.013	0.059	0.009
$u0_j$, Variance	0.033**	0.028**	0.050**	0.040**	0.026**
$u00_j$, Variance	0.029**	0.034**	0.032**	0.039**	0.027**
Intercept, γ00, reliability	0.527	0.469	0.552	0.510	0.452
Intercept, γ000, reliability	0.620	0.664	0.557	0.638	0.614

+ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 10E: Impact of Treatment on Students' Writing Ability for Teacher Experience, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Total Score	Word Choice	Content Score	Organization Score	Conventions Score
Intercept, γ000	2.361**	2.305**	2.461**	2.351**	2.256**
TREATMT, γ010	0.041	0.067	0.049	0.038	0.033
COHORT2, γ020	-0.040	-0.033	-0.005	0.004	-0.114+
GRADE4, γ030	0.077	0.103*	0.103+	0.095+	0.104*
GRADE5, γ040	0.141**	0.163**	0.227**	0.167**	0.158*
TEXP, γ050	0.007*	0.007**	0.008*	0.007*	0.006*
TRT*TEXP, γ060	-0.004	-0.003	-0.006	-0.005	-0.003
FALL, γ100	0.432**	0.375**	0.323**	0.358**	0.469**
MALE, γ200	-0.164**	-0.168**	-0.200**	-0.186**	-0.164**
BLACK, γ300	-0.182**	-0.212**	-0.202**	-0.179**	-0.154**
HISPANIC, γ400	-0.174**	-0.200**	-0.192**	-0.196**	-0.123*
OTHER, γ500	-0.014	-0.008	-0.044	-0.036	-0.004
$u0_j$, Variance	0.032**	0.026**	0.048**	0.039**	0.025**
$u00_j$, Variance	0.028**	0.031**	0.030**	0.037**	0.025**
Intercept, γ00, reliability	0.515	0.449	0.541	0.507	0.439
Intercept, γ000, reliability	0.616	0.653	0.547	0.626	0.607

+ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 10F: Impact of Treatment on Students' Writing Ability for <u>Teacher Confidence</u>, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Total Score	Word Choice	Content Score	Organization Score	Conventions Score
Intercept, γ000	2.403**	2.315**	2.492**	2.371**	2.352**
TREATMT, γ010	0.019	0.070	0.046	0.005	-0.031
COHORT2, γ020	-0.039	-0.029	-0.003	0.010	-0.122^{+}
GRADE4, γ030	0.068	0.094^{+}	0.095	0.085	0.096*
GRADE5, γ040	0.163**	0.184**	0.253**	0.195**	0.180*
TCONF, γ050	0.008	0.020	0.015	0.016	-0.010
TRT*TCONF, γ060	-0.004	-0.010	-0.017	-0.005	0.013
FALL, γ100	0.432**	0.376**	0.323**	0.359**	0.469
MALE, γ200	-0.162**	-0.166**	-0.199**	-0.185**	-0.161**
BLACK, γ300	-0.183**	-0.212**	-0.205**	-0.180**	-0.155**
HISPANIC, γ400	-0.177**	-0.204**	-0.195**	-0.199**	-0.126*
OTHER, γ500	-0.012	-0.009	-0.044	-0.038	0.006
$u0_j$, Variance	0.033**	0.028**	0.051**	0.041**	0.026**
$u00_j$, Variance	0.029**	0.031**	0.030**	0.036**	0.028**
Intercept, γ00, reliability	0.529	0.471	0.555	0.517	0.446
Intercept, γ000, reliability	0.613	0.643	0.535	0.616	0.627

$^{+}$ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 11A: Impact of Treatment on Students' Writing Rating and Attitudes for Gender, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Teacher Rating	Writing Attitudes Scale	In School Writing Scale	Writing Grades Scale
Intercept, γ000	2.666**	24.369**	11.755**	4.500**
TREATMT, γ010	0.111	-0.131	-0.705**	-0.084
COHORT2, γ020	0.034	-0.288	0.258	0.034
GRADE4, γ030	0.146*	0.232	-0.143	-0.055
GRADE5, γ040	0.157	0.739*	0.428	0.012
FALL, γ100	0.638**	0.236**	0.222**	0.196**
MALE, γ200	-0.161**	1.680**	0.369**	0.410**
TRT*MALE, γ_{210}	0.004	-0.169	0.116	**-0.200**$^{+}$
BLACK, γ300	-0.169**	-0.751**	0.078	-0.148
HISPANIC, γ400	-0.173**	0.002	0.225	-0.008
OTHER, γ500	-0.050	0.507	0.579**	0.194*
$u0_j$, Variance	0.097**	0.448**	0.614**	0.048**
$u00_j$, Variance	0.006	0.503**	0.131*	0.001
Intercept, γ00, reliability	0.702	0.297	0.653	0.351
Intercept, γ000, reliability	0.129	0.537	0.336	0.028

$^{+}$ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 11B: Impact of Treatment on Students' Writing Rating and Attitudes for 4th Grade, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Teacher Rating	Writing Attitudes Scale	In School Writing Scale	Writing Grades Scale
Intercept, γ000	2.650**	24.276**	11.675**	4.480**
TREATMT, γ010	0.141	0.042	-0.550**	-0.049
COHORT2, γ020	0.035	-0.282	0.264	0.035
GRADE4, γ030	0.180^{+}	0.433	0.038	-0.014
GRADE5, γ040	0.156^{+}	0.745*	0.428	0.016
TRT*GRADE4, γ050	-0.071	-0.401	-0.368	-0.077
FALL, γ100	0.638**	0.236**	0.222**	0.196**
MALE, γ200	-0.159**	1.595**	0.426**	0.311**
BLACK, γ300	-0.168**	-0.742*	0.080	-0.144
HISPANIC, γ400	-0.172**	0.008	0.228	-0.008
OTHER, γ500	-0.049	0.521	0.583**	0.199*
$u0_j$, Variance	0.097**	0.433**	0.603**	0.048**
$u00_j$, Variance	0.006	0.510**	0.134*	0.001
Intercept, γ00, reliability	0.701	0.290	0.649	0.352
Intercept, γ000, reliability	0.136	0.542	0.343	0.024

$^{+}$ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 11C: Impact of Treatment on Students' Writing Rating and Attitudes for 5th Grade, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Teacher Rating	Writing Attitudes Scale	In School Writing Scale	Writing Grades Scale
Intercept, γ000	2.676**	24.361**	11.754**	4.503**
TREATMT, γ010	0.091	-0.119	-0.702**	-0.092
COHORT2, γ020	0.035	-0.289	0.259	0.034
GRADE4, γ030	0.145*	0.234	-0.144	-0.053
GRADE5, γ040	0.049	0.792*	0.448*	-0.033
TRT*GRADE5, γ050	0.200	-0.102	-0.042	0.096
FALL, γ100	0.638**	0.236**	0.222**	0.196**
MALE, γ200	-0.158**	1.595**	0.427**	0.312**
BLACK, γ300	-0.169**	-0.748**	0.076	-0.146
HISPANIC, γ400	-0.173**	0.001	0.225	-0.009
OTHER, γ500	-0.050	0.511	0.578**	0.198*
$u0_j$, Variance	0.095**	0.450**	0.613**	0.049**
$u00_j$, Variance	0.007	0.504**	0.130*	0.001
Intercept, γ00, Reliability	0.699	0.298	0.653	0.353
Intercept, γ000, Reliability	0.145	0.537	0.335	0.024

$^{+}$ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 11D: Impact of Treatment on Students' Writing Rating and Attitudes for Low Fall Essay Scores, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Teacher Rating	Writing Attitudes Scale	In School Writing Scale	Writing Grades Scale
Intercept, γ000	2.707**	24.360**	11.718**	4.500**
TREATMT, γ010	0.105	-0.131	-0.702**	-0.083
COHORT2, γ020	0.031	-0.289	0.258	0.034
GRADE4, γ030	0.076	0.245	-0.078	-0.056
GRADE5, γ040	0.052	0.770*	0.525^{+}	0.014
FALL, γ100	0.577**	0.236**	0.221**	0.197**
MALE, γ200	-0.137**	1.587**	0.392**	0.312**
BLACK, γ300	-0.163**	-0.756**	0.070	-0.147
HISPANIC, γ400	-0.168**	-0.005	0.220	-0.009
OTHER, γ500	-0.023	0.497	0.585**	0.194*
FallLOW, γ600	-0.327**	0.176	0.092	0.039
TRT*FallLOW, γ_{610}	-0.005	-0.174	**0.370**$^{+}$	-0.083
$u0_j$, Variance	0.097**	0.455**	0.601**	0.049**
$u00_j$, Variance	0.006	0.502**	0.133*	0.001
Intercept, γ00, Reliability	0.709	0.301	0.649	0.355
Intercept, γ000, Reliability	0.128	0.535	0.342	0.023

$^{+}$ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 11E: Impact of Treatment on Students' Writing Rating and Attitudes for <u>Teacher Experience</u>, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Teacher Rating	Writing Attitudes Scale	In School Writing Scale	Writing Grades Scale
Intercept, γ000	2.684**	24.098**	11.459**	4.531**
TREATMT, γ010	0.007	0.121	-0.401^{+}	-0.147
COHORT2, γ020	0.048	-0.290	0.241	0.042
GRADE4, γ030	0.141*	0.264	-0.107	-0.057
GRADE5, γ040	0.185*	0.624*	0.276	0.039
TEXP, γ050	-0.002	0.027	0.031*	-0.003
TRT*TEXP, γ060	0.010	-0.026	-0.031	0.006
FALL, γ100	0.638**	0.236**	0.224**	0.196**
MALE, γ200	-0.159**	1.583**	0.419**	0.313**
BLACK, γ300	-0.171**	-0.729**	0.098	-0.151^{+}
HISPANIC, γ400	-0.173**	0.027	0.238	-0.012
OTHER, γ500	-0.054	0.480	0.574**	0.193*
$u0_j$, Variance	0.093**	0.443**	0.558**	0.048**
u_{00j}, Variance	0.007	0.446**	0.150**	0.001
Intercept, γ00, Reliability	0.694	0.295	0.632	0.352
Intercept, γ000, Reliability	0.156	0.508	0.379	0.016

$^{+}$ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.

Exhibit 11F: Impact of Treatment on Students' Writing Rating and Attitudes for <u>Teacher Confidence</u>, Combined Cohort 1 and 2 Impact Sample, Spring 2006 and Spring 2007 (Weighted for Non-Response)

Variables	Teacher Rating	Writing Attitudes Scale	In School Writing Scale	Writing Grades Scale
Intercept, γ000	2.680**	24.366**	11.922	4.614**
TREATMT, γ010	0.220^{+}	-0.028	-0.623^{+}	-0.132
COHORT2, γ020	0.020	-0.300	0.213	0.020
GRADE4, γ030	0.155*	0.245	-0.119	-0.053
GRADE5, γ040	0.142	0.730*	0.415	0.017
TCONF, γ050	-0.003	0.001	-0.055	-0.039
TRT*TCONF, γ060	-0.040	-0.038	-0.034	0.017
FALL, γ100	0.638**	0.236**	0.222**	0.196**
MALE, γ200	-0.158**	1.596**	0.429**	0.314**
BLACK, γ300	-0.175**	-0.754**	0.064	-0.155
HISPANIC, γ400	-0.169**	0.007	0.239	-0.001
OTHER, γ500	-0.042	0.517	0.600**	0.204*
$u0_j$, Variance	0.094**	0.449**	0.609**	0.046**
$u00_j$, Variance	0.007	0.504**	0.120*	0.001
Intercept, γ00, Reliability	0.696	0.297	0.651	0.335
Intercept, γ000, Reliability	0.162	0.537	0.319	0.048

$^{+}$ Statistical trend at the 90 percent confidence level.

* Statistically significant at the 95 percent confidence level.

Chapter 5: Discussion

Summary of Findings

Overall, there were no statistically significant impacts on student's writing ability, and a similar lack of statistically significant impacts on teacher ratings of their student's writing ability. However, student responses to the survey indicated statistically significant differences on several individual questionnaire items, and the multivariate analysis confirms that *Writing Wings* appears to have had a relatively large positive impact on the frequency of student-reported in-school writing.

Looking at the possibility of impacts on selected subgroups of students, these more exploratory analyses indicate that there may be an impact on student's written organization skills for 5th grade students, and at least a suggestion of a positive impact on the frequency of in-school writing for those students who had relatively low initial writing scores.

Why are Impacts Weak?

These results are disappointing but may not be surprising in light of several conditions may have contributed to the observed lack of impacts on student outcomes:

- First, teachers were reportedly at a fairly high level of writing instruction at the start of the study. Nearly seven out of ten teachers at the time of the baseline survey reportedly taught writing three days/week, and about three-quarters reportedly spent 30-60 minutes on each writing lesson. Moreover, teachers initially reported that they frequently taught important writing skills including prewriting (e.g., using webs to organize one's thoughts), preparing drafts, and editing and revising one's work. Aspects of the writing process that were less frequently incorporated at the time of the baseline survey included having students revise their drafts based on feedback, publishing their work, sharing drafts with other students, and working with a partner on developing their written piece.

  Treatment group teachers were found to have modified their instructional practices as a consequence of *Writing Wings* − they reportedly taught writing more often and for longer sessions, had students execute multiple aspects of the writing process more frequently, and felt more confident teaching writing – but these intermediate impacts on teachers may have been too small to have a large enough subsequent impact on student outcomes to be detected with this ample size.

- Second, the initial training of teachers was delayed for Cohort 1. Because of the late grant award– and the associated effect on school recruitment, baseline data collection, and random assignment – 13 of the 17 schools in Cohort 1 did not receive teacher training until after October 15th with seven schools being trained between November 15th and the 30th. For Cohort 2, 16 of the 22 schools were trained by September 30th and only 2 schools were delayed until early November.

- Third, the time required for teachers to reach an acceptable level of implementation was slower than originally expected. As discussed in Chapter 2, approximately two

months after their initial training, only about half of the treatment group teachers had reached at least the "routine" implementation level and this increased to about 70 percent after five months. It takes time to implement a new and complicated instructional program like *Writing Wings*, particularly when a substantial change in teacher practices may be needed to create a meaningful impact on student's writing ability. Consequently, the expectation that a single year of experience with the new curriculum would be sufficient to reach the required level of teacher proficiency may have been incorrect. In retrospect, it may have been better to allow teachers a year to learn the instructional program and to assess the impact on students in the second year following the start of teacher training.

- Fourth, the level of teacher training and support may have been inadequate to create the necessary change in the quantity and quality of writing instruction that students received. That is, the one-day initial training, followed by an average of three support visits in September-December, January-March, and April-May may not have been an adequate level of teacher professional development and support.
- Finally, students' initial writing performance was generally low; students scored, on average, a "2" or "3" on the 5-point essay scale at the time of the baseline assessment. Thus, teachers in the current study were instructing many students with substantial need in terms of writing skills. Moreover, student ability influences teachers' instruction, as teachers mentioned, making instruction challenging when there is a range of student writing ability to deal with in their classrooms. However, when teachers are utilizing classroom resources to the best of their ability and full implementation of the treatment program is not achieved, minimal treatment impacts on students' writing performance are not surprising, even for low-achieving students.

Suggestions for the Future

Writing Wings is a complicated program that requires a significant amount of effort on the part of classroom teachers to reach a desirable level of instructional practice. The professional development materials alone consist of two manuals – a Teacher's Manual that contains detailed lesson guides that extends to over 400 pages, and a manual on Language Mechanics that provides instruction guides, student worksheets and assessments focusing on grammar and writing mechanics (consisting of another 350 pages). Not only is this a lot for teachers to absorb and to figure out how to best integrate the instruction into their existing classroom routines, but early feedback from teachers indicates that they often struggled with finding the added time needed to bring this level of writing instruction into a schedule that is already stretched.

As a consequence, SFAF should take these results as an opportunity to re-evaluate their instructional program, and the associated professional development model, to find a way to create the level of change in instructional practice needed to improve student's writing ability. This may involve a simplification of the manuals and lesson plans to make it easier for teachers to "get up to speed" quickly, a greater focus on helping teachers better integrate writing into all of their subject area instruction (i.e., expanding the amount of instruction), and a more intensive professional development model.

In addition, in light of the relatively low level of student writing ability that was observed in the Fall, SFAF should consider exposing students to the program for consecutive years (e.g., 3rd and 4th grades), and possibly integrating the instruction across subject areas, to get the desired improvement on student achievement.

This is not to say that impacts would have been different if the program and the training had been implemented differently, but at least these suggested changes would remove some of the possible reasons why the estimated effects were so weak thereby increasing our confidence in any observed impacts on student outcomes.

Appendices

A. Site Agreement and Parent Consent Form.
B. Student and Teacher Questionnaires.
C. Impacts on Teacher Outcomes.

Appendix A-1: District, School, and Researcher Agreement Form
The *Writing Wings* Study

The following is an agreement between the _____ School District and Chesapeake Research Associates, LLC (CRA). This agreement applies to the research being conducted by CRA as part of the *Evaluation of "Writing Wings": Writing Instruction for Disadvantaged Elementary School Children.* This project is funded by the U.S. Department of Education, Institute of Education Sciences, and extends from July 1, 2005 to August 1, 2007.

FOR _____ SCHOOL(S):

Writing Wings is a curriculum and support package that requires a minimum of 40 minutes of instruction, four to five days a week.

Before the school year begins, at each participating school, CRA will assign one third-grade teacher and one fourth-grade teacher to implement Writing Wings and one third-grade teacher and one fourth-grade teacher to serve as the "comparison" or "non-Writing Wings" teacher by not implementing Writing Wings. This selection **will be done at random** and will be done **after** the school has assigned students to teachers.

The school(s) agrees to provide the following to researchers from CRA:

1. Participation of teachers and students in two third-grade and two fourth-grade classrooms per school.
2. Assurance that
    - Once assigned to implement Writing Wings, teachers will make their best effort to fully implement and maintain facilitation of Writing Wings for the duration of the 2006-07 school year.
    - Once assigned to serve as the "comparison" or "non-Writing Wings" teacher, comparison teachers will **not** implement any part Writing Wings for the duration of the 2006-07 school year.
    - All school staff assigned to implement Writing Wings, and all school staff whose classrooms are assigned to serve as the comparison classrooms, will allow visits by trainers to observe their classrooms.
    - All participating school staff will monitor the receipt and storage of curriculum materials.

3. Copies of student rosters including: student names, student identification numbers, ethnicity, gender, and name of homeroom teacher.

4. Permission to collect the following data at the beginning and end of the 2006-07 school year:
    - student writing assessment
    - a brief student questionnaire
    - a brief teacher questionnaire

    Data collection will occur at the start of the fall semester in 2006, and toward the end of the spring semester in 2007. The duration of collection is estimated at one hour in each participating classroom. (Note: at the discretion of the school, all students can be tested at the same time.)

FOR CHESAPEAKE RESEARCH ASSOCIATES, LLC:

Chesapeake Research Associates, in conjunction with the Success for All Foundation, will provide the following to participating schools at no cost:

1. Writing Wings for classrooms whose teachers are assigned to implement Writing Wings in the 2006-07 school year (see table below), including the following:
   - One day of initial training, plus follow-up support visits, e-mails, and phone consultations.
   - Manuals, trade books, and other classroom materials needed for use of the program.
2. Writing Wings for the non-Writing Wings third- and fourth-grade classes in the 2007-08 school year (see table below), including:
   - One day of initial training.
   - Manuals and trade books.

Progression of *Writing Wings Evaluation*: Example of a School with Two Classes per Grade

	2006-07		2007-08
Grade 3	1 Writing Wings class/ 1 non-Writing Wings class	→	2 Writing Wings classes
Grade 4	1 Writing Wings class/ 1 non-Writing Wings class	→	2 Writing Wings classes

3. Waiver of the "Experienced Sites" conference fees for two Writing Wings teachers for the 2006-07 and 2007-08 school years.

Confidentiality

Information we collect will be kept strictly confidential to the full extent allowed by law. The results will be reported only in group form, such as "70 percent of third graders could write at grade level." We will not present student achievement data by teacher, by class, or by school. In this way, the confidentiality of students, teachers, and schools will be carefully guarded.

District Official:

_____
Signature

_____
Title

_____
Date

School Principal:

_____
Signature

_____
Date

Chesapeake Research Associates, LLC

Michael J. Puma, President

_____
Signature

_____
Date

Appendix A-2: CONSENT TO PARTICIPATE IN RESEARCH

PROJECT TITLE: ***An Evaluation of "Writing Wings": Writing Instruction for Disadvantaged Elementary School Children***

PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR: ***Michael J. Puma, 410-897-4968***

SPONSOR: ***US Department of Education, Institute for Education Sciences***

The Georgetown University Institutional Review Board (IRB) has approved this research project. For information on your rights as a research subject, call the Institutional Review Board office at 202-687-1506.

INTRODUCTION

You are invited to consider participating in a research study to investigate the effectiveness of *Writing Wings*, an instructional program for elementary school children. This form will describe the purpose and nature of the research, its possible risks and benefits, and your rights as a participant in the study. The decision to participate, or not to participate, is yours. If you decide to participate, please be sure to sign and date the last page of this form.

WHY IS THIS RESEARCH STUDY BEING DONE?

In this research study, we are evaluating *Writing Wings* to determine how well children learn to write when teachers are well trained and use selected instructional materials. Writing is critically important for children's success in school and in later life.

HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?

Nationally, about 3,000 students will take part in this study.

WHAT IS INVOLVED IN THE STUDY?

Your child's school has decided to try a new writing program – called *Writing Wings* -- but for this year, only half the 3rd and 4th grade classes (and in some schools, the 5th grade classes) will be able to participate. Consequently, classes will be picked by "lottery" to receive the new instruction this year with the other classes receiving the new instruction next year. A computer will determine your child's group through a process that is much like picking names out of a hat. This process is called randomization. Your child's chance of being in any group is one in two.

To determine the effectiveness of the new program, your child will be given a writing test in the Fall and again in Spring along with a brief questionnaire about their writing. Completion of both the test and the survey will require a total of about 1 hour and will be administered in your child's classroom.

HOW LONG WILL I BE IN THE STUDY?

We expect that your child will be in the study for just 1 year.

The investigators or sponsors may stop the study at any time they judge it is in your best interest or for a variety of other reasons. They can do this without your consent. You can stop participating at any time. However, if you decide to stop participating in the study, we encourage you to talk to the researcher first.

WHAT ARE THE RISKS OF THE STUDY?

This study involves low risk to your child limited to the inadvertent disclosure of his/her performance on the writing test.

ARE THERE BENEFITS TO TAKING PART IN THE STUDY?

Participating schools will receive the *Writing Wings* instructional program – including teacher training, books, and teaching materials – **at no cost.**

It is reasonable to expect that children who receive the *Writing Wings* instruction will improve their writing skills. However, we cannot guarantee that your child will personally experience benefits from participating in this study. Others may benefit in the future from the information we obtain in this study.

WHAT ABOUT CONFIDENTIALITY?

Your child's name will not be used when data from this study are published.

Every effort will be made to keep your child's research records and other personal information confidential. However, we cannot guarantee absolute confidentiality.

We will take the following steps to keep information about you confidential, and to protect it from unauthorized disclosure, tampering, or damage:

1. All staff and field workers will sign an assurance of confidentiality pledging to keep completely confidential the names of respondents and all information that is collected.
2. An employee, upon encountering a respondent or information pertaining to a respondent that s/he knows personally, will immediately terminate the activity and contact her/his supervisor for instructions.
3. Survey data containing personal identifiers will be kept in a locked container or a locked room when not being used each working day in routine survey activities.
4. Serial numbers will be assigned to respondents prior to creating a machine-processible record and identifiers such as name and address will not be a part of the machine record.

WILL I BE PAID FOR PARTICIPATING?

Students will **not** be paid for participating in this study. But, as noted above, participating schools will receive the *Writing Wings* instructional program – including teacher training, books, and teaching materials – at no cost.

WHAT ARE MY RIGHTS AS A RESEARCH PARTICIPANT?

Participation in this study is entirely voluntary at all times. You have the right not to participate at all or to leave the study at any time. Deciding not to participate or choosing to leave the study will not result in any penalty or loss of benefits to which you are entitled. If you decide to leave the study, the procedure is to contact your child's school principal.

WHOM DO I CONTACT IF I HAVE QUESTIONS OR PROBLEMS?

Call Michael J. Puma at (410) 897-4968 during regular Eastern Time business hours if you have questions about the study, any problems, unexpected physical or psychological discomforts, any injuries, or think that something unusual or unexpected is happening.

Permission for a Child to Participate in Research

As parent or legal guardian, I authorize _____ (child's name)
to become a participant in the research study described in this form.

_____ _____

Parent or Legal Guardian's Signature Date

Appendix B: Questionnaires

School Name: _____
Teacher Name: _____

The *Writing Wings* Project: Teacher Questionnaire

Spring 2006

Part 1. Your Professional Background

Q1. What is your highest degree? (✓ Check one)

a. BA or BS ❑1

b. MA or MS ❑2

c. PhD or EdD ❑3

d. Other (Please describe) _____ ❑4

Q2. Which of the following statements most accurately describes the type of teaching credential that you currently hold? (✓ Check one)

a. A **regular** or **standard state certificate** ..... ❑1

b. An **emergency certificate** or **waiver** that is issued for a specified time period to persons with insufficient teacher preparation ..... ❑2

c. Other (Please describe _____) ..... ❑3

d. I am not certified ..... ❑4

Q3. How many years have you been a full-time classroom teacher?

_____ (number of years)

Part 2. Your Writing Instruction

Q4. Generally speaking, how often do you teach writing? (✓ Check one)

a. Rarely ❑1

b. 1-2 days/week ❑2

c. 3-4 days/week ❑3

d. Every day ❑4

Q5. On the days you teach writing, about how much time do you spend (including instruction time and time on student activities)? (✓ Check one)

a. Less than 30 minutes ❑1

b. Between 30 minutes and 1 hour ❑2

c. More than 1 hour ❑3

Q6. How do you grade your student's writing assignments?

a. Single grade for whole assignment ❑1
b. Separate grade for content and mechanics ❑2
c. Other (Specify) _____ ❑3

Q7. What do you focus on most in assigning a grade?

		Very Important	Important	Not Important
a.	Complete response to assignment	❑1	❑2	❑3
b.	Good Organization	❑1	❑2	❑3
c.	Use of interesting language	❑1	❑2	❑3
d.	Punctuation	❑1	❑2	❑3
e	Spelling	❑1	❑2	❑3
f.	Complete Sentences	❑1	❑2	❑3
g.	Neatness	❑1	❑2	❑3

Q8. On the days you **teach** writing, about how often do you spend time on the following activities? (✓ Check one per row)

Activity	Every lesson (=4)	Most lessons (=3)	Some lessons (=2)	Never (=1)
Prewriting skills including: (1) the generation and organization of ideas (brainstorming), (2) considering the topic, audience, purpose, and form, of the writing, and (3) talk-write with a partner.				
Drafting skills focusing on ideas, message, and meaning over mechanics, such as spelling and punctuation.				
Sharing, learning how to share, give positive and constructive feedback, and how to make changes based on feedback.				
Revising skills, learning how to add description, elaborate on a topic, clarify information, and refine vocabulary.				
Editing skills, a final check on spelling, punctuation, and grammar.				
Publishing, sharing the completed writing with others.				
Worksheet activities on specific grammar skills				

Q9. When your students work on their writing, about how often do they do each of the following activities? (✓ Check one)

		Always	Sometimes	Never
a.	Work on development with a partner	❑1	❑2	❑3
b.	Use graphic organizers	❑1	❑2	❑3
c.	Write a first draft	❑1	❑2	❑3
d.	Revise their drafts	❑1	❑2	❑3
e	Share drafts with other students	❑1	❑2	❑3
f.	Revise draft based on feedback	❑1	❑2	❑3
g.	Proofread and make final edits	❑1	❑2	❑3
h.	Publish their work	❑1	❑2	❑3

Q10. Do you incorporate writing activities into the following subjects?

		YES	NO
a.	Social studies	❑1	❑2
b.	Math	❑1	❑2
c.	Science	❑1	❑2
d.	Arts	❑1	❑2
e	Other (please specify) _____		❑1

Q11. Do your students keep a portfolio of their writing?

a. Yes ❑1

b. No ❑2

Q12. What do you find the most difficult part of writing instruction? _____

_____

_____

Q.13 How comfortable do you feel teaching writing? (✓ Check one)

a. Not at all confident ❑[1]

b. Somewhat uncertain ❑[2]

c. Comfortable ❑[3]

d. Pretty confident ❑[4]

e. Very confident ❑[5]

Q14. Which of the following activities affect your time for writing instruction? (Check all that apply)

a. Field trips ❑[1]

b. Testing (including State assessments) ❑[2]

c. Test preparation ❑[3]

d. Other curricula demands ❑[4]

e. Schedule changes for special school programs ❑[5]

f. Other_____ ❑[6]

Thank you for your time and assistance!

PRINT YOUR NAME: _____

YOUR TEACHER'S NAME: _____

The *Writing Wings* Project: Student Questionnaire

Spring 2006

Q1. How often do you write for fun? (✓ Check one)

Every day..... ❑[1]
Most days ..... ❑[2]
Some days..... ❑[3]
Never ..... ❑[4]

Q2. What do you think about writing?

2A. Writing helps me think more clearly. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

2B. Writing helps me tell others what I think. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

2C. Writing helps me tell others how I feel. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

2D. Writing helps me understand my own feelings. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

2E. People who write well do better in school. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

Q3. Tell us about your writing.

3A. I like to write. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

3B. I am a good writer. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

3C. People like what I write. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

3D. I write on my own outside of school. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

3E. I don't like to write things that will be graded. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

3F. If I didn't have to write for school, I wouldn't write anything. (✓ Check one)

Really True ..... ❑[1]
True ..... ❑[2]
Not True..... ❑[3]
Really Not True..... ❑[4]

Q4. When you do writing in class, how often do you usually do each of these things?

4A. Use a chart or a web to organize your thoughts. (✓ Check one)

Most of the time. .... ❑[1]
Sometimes..... ❑[2]
Never ..... ❑[3]

4B. Write one or more drafts. (✓ Check one)

Most of the time. .... ❑[1]
Sometimes..... ❑[2]
Never ..... ❑[3]

4C. Share what you write with a partner. (✓ Check one)

Most of the time. .... ❑[1]
Sometimes..... ❑[2]
Never ..... ❑[3]

4D. Revise your writing based on the suggestions of your teacher or another student. (✓ Check one)

Most of the time. .... ❑[1]
Sometimes..... ❑[2]
Never ..... ❑[3]

4E. Do a final check of your spelling and punctuation. (✓ Check one)

Most of the time. .... ❑[1]
Sometimes..... ❑[2]
Never ..... ❑[3]

4F. Share your completed work with others. (✓ Check one)

Most of the time. .... ❑[1]
Sometimes..... ❑[2]
Never ..... ❑[3]

4G. Practice punctuation or grammar in class. (✓ Check one)

Most of the time. .... ❑[1]
Sometimes..... ❑[2]
Never ..... ❑[3]

Q5. How do you get a good grade in a writing assignment?

5A. Write an interesting paper. (✓ Check one)

Very important..... ❑[1]
Important..... ❑[2]
Not important ..... ❑[3]

5B. Make few spelling errors. (✓ Check one)

Very important..... ❑[1]
Important..... ❑[2]
Not important ..... ❑[3]

5C. Write neatly. (✓ Check one)

Very important..... ❑[1]
Important..... ❑[2]
Not important ..... ❑[3]

Q6. How many days do you usually get homework? (✓ Check one)

Don't get homework ..... ❑[0]
1 day per week..... ❑[1]
2 days per week..... ❑[2]
3 days per week..... ❑[3]
4 days per week..... ❑[4]
Every day..... ❑[5]

Q7. How often do you have writing to do for homework?

More than once per week…… ❑[0]
Once per week……..❑[1]
More than once per month……. ❑[2]
About once per month.……❑[3]
Rarely or never……..❑[4]

Q8. How many minutes do you usually spend doing homework per day?

_____ (minutes)

Appendix C: Impacts on Teachers – Mean Differences, Spring Teacher Survey Combined Cohort 1 and 2 Study Samples

Teacher Question	*Treatment Group*	*Control Group*	*Difference*
Generally speaking, how often do you teach writing? (1=Rarely, 4=Every day)	3.21	2.93	**0.28***
On the days you teach writing, about how much time do you spend (including instruction time and time on student activities)? (1=Less than 30 minutes, 3=More Than 1 Hour)	2.03	1.85	0.18*
On the days you teach writing, about how often do you spend time on the following activities?			
▪ **Prewriting skills** including: (1) the generation and organization of ideas (brainstorming), (2) considering the topic, audience, purpose, and form, of the writing, and (3) talk-write with a partner.	3.43	3.23	0.20
▪ **Drafting skills** focusing on ideas, message, and meaning over mechanics, such as spelling and punctuation.	3.36	3.21	0.15
▪ **Sharing**, learning how to share, give positive and constructive feedback, and how to make changes based on feedback.	2.99	2.59	**0.40****
▪ **Revising skills**, learning how to add description, elaborate on a topic, clarify information, and refine vocabulary.	3.00	2.81	0.19
▪ **Editing skills**, a final check on spelling, punctuation, and grammar.	3.08	2.92	0.16
▪ **Publishing**, sharing the completed writing with others.	2.89	2.54	**0.35****
▪ Worksheet activities on specific **grammar skills.**	2.53	2.24	**0.29***
When your students work on their writing, about how often do they do each of the following activities? (1=Always, 3=Never)			
▪ Work on development with a partner.	1.48	1.92	**-0.44****
▪ Use graphic organizers	1.13	1.40	**-0.27****
▪ Write a first draft	1.08	1.17	-0.09
▪ Revise their drafts	1.20	1.21	-0.01
▪ Share drafts with other students	1.32	1.87	**-0.55****
▪ Revise draft based on feedback	1.35	1.64	**-0.29****
▪ Proofread and make final edits	1.36	1.37	-0.01
▪ Publish their work	1.48	1.68	**-0.20***
How comfortable do you feel teaching writing? (1=Not At All Confident, 5=Very Confident)	3.65	3.26	**0.39***

* Statistically significant at the 95 percent confidence level.

** Statistically significant at the 99 percent confidence level.