

A Meta-Analysis of Single Subject Design Writing Intervention Research

Leslie Ann Rogers and Steve Graham
Vanderbilt University

There is considerable concern that students do not develop the writing skills needed for school, occupational, or personal success. A frequent explanation for this is that schools do not do a good job of teaching this complex skill. A recent meta-analysis of true- and quasi-experimental writing intervention research (S. Graham & D. Perin, 2007a) addressed this issue by identifying effective instructional writing practices. The current review extends this earlier work by conducting a meta-analysis of single subject design writing intervention studies. The authors located 88 single subject design studies where it was possible to calculate an effect size. They calculated an average effect size for treatments that were tested in 4 or more studies, using a similar outcome measure in each study. This resulted in the identification of 9 writing treatments that were supported as effective. These were strategy instruction for planning/drafting, teaching grammar and usage, goal setting for productivity, strategy instruction for editing, writing with a word processor, reinforcing specific writing outcomes, use of prewriting activities, teaching sentence construction skills, and strategy instruction for paragraph writing.

Keywords: writing, composition, meta-analysis, instruction, single subject design

According to the most recent National Assessment of Educational Progress (Persky, Daane, & Jin, 2003), many children do not learn to write well enough to meet classroom writing demands. The writing of 2 out of every 3 students in Grades 4, 8, and 12 was below grade-level proficiency. Concerns about writing are not limited to elementary and secondary schools, however, as college instructors estimated that 50% of high school graduates are not prepared for college-level writing demands (Achieve, Inc., 2005). Moreover, American businesses spend \$3.1 billion annually for writing remediation (National Commission on Writing, 2004). In 2003, these problems led the National Commission on Writing to conclude that the writing of students in the United States “is not what it should be” (p. 7).

Students who do not learn to write well are at a disadvantage. At school, weaker writers are less likely than their more skilled peers to use writing to support and extend learning in content classrooms (Graham & Perin, 2007a). Their grades are more likely to suffer, particularly in classes where writing is the main tool (e.g., via tests and papers) for assessing progress (Graham, 2006b). Their chances of attending college are reduced, as writing is now used to evaluate many applicants’ qualifications. At work, writing has become a gateway for employment and promotion (see reports by the National Commission on Writing, 2004, 2005). Employees in business and government, for instance, are expected to produce written documentation, visual presentations, memoranda, technical re-

ports, and electronic messages. Socially, adults who do not write well may not be able to participate fully in civic life, as e-mail and text messaging have progressively supplanted telephones as a primary means for communicating. On a personal level, people use writing to explore who they are, to combat loneliness, and to chronicle their experiences. Writing about one’s feelings and experiences, for example, is beneficial psychologically and physiologically (see Smyth, 1998, for a meta-analysis of this research).

Why do so many students not write well enough to meet grade level demands? One possible reason is that schools do not do an adequate job of teaching this complex skill. The National Commission on Writing (2003) charged that this is the most neglected of the three *Rs* in the American classroom and offered the following recommendations: double time students spend writing, assess their writing progress, use technology to advance the learning and teaching of writing, and better prepare teachers to teach writing. The impact of these recommendations is likely to be reduced if teachers do not use effective instructional practices.

Need for a Meta-Analysis of Writing Interventions Tested Via Single Subject Design Studies

A useful approach for identifying effective writing practices is to conduct systematic reviews of writing intervention research. Since the mid-1980s there have been a number of meta-analyses of the writing intervention literature (e.g., Hillocks, 1986; Graham & Perin, 2007a, 2007b). With meta-analysis, an effect size is computed for each empirical study investigating a specific treatment and then is averaged across studies to provide a summary statistic on the intervention’s effectiveness (Lipsey & Wilson, 2001). Most of the meta-analyses of writing intervention research have focused on treatments tested in true- and quasi-experimental studies (Bangert-Drowns, 1993; Bangert-Drowns, Hurley, & Wilkinson, 2004; Goldring, Russell, & Cook, 2003; Graham, 2006a; Graham & Harris, 2003; Graham & Perin, 2007a, 2007b; Hillocks, 1986).

Editor’s Note. Tom Scruggs served as the action editor for this article.—KRH

Leslie Ann Rogers and Steve Graham, Department of Special Education, Vanderbilt University.

Correspondence concerning this article should be addressed to Steve Graham, Peabody 328, 230 Appleton Place, Vanderbilt University, Nashville, TN 37203. E-mail: steve.graham@vanderbilt.edu

However, two of the meta-analyses also computed effect sizes for treatments tested via single subject designs (Graham, 2006a; Graham & Harris, 2003). These two reviews only examined a single treatment: teaching strategies for planning/revising. Consequently, the primary purpose of this article was to conduct a more extensive meta-analysis of single subject design writing interventions in order to identify effective writing practices for students in Grades 1–12.

Like true-experimental studies, single subject designs can be used to test whether a treatment is responsible for observed changes in performance. As Horner et al. (2005) noted, “Single-subject design is experimental rather than correlational or descriptive, and its purpose is to document causal or functional relationships between independent and dependent variables” (p. 166). Major threats to internal validity are controlled through within- and between-subjects comparisons, and external validity is enhanced through systematic replication. In single subject design studies, each participant serves as her/his own control, with performance prior to as well as during and/or after intervention repeatedly measured to establish performance patterns before treatment and comparison of performance patterns across experimental phases (e.g., baseline versus treatment). To establish experimental control, the independent variable or treatment is actively manipulated to determine its effects on the dependent measure(s) (Horner et al., 2005).

One manipulation (reversal) for establishing experimental control involves introduction and withdrawal of the treatment (e.g., a stable baseline pattern of performance is established, followed by introduction of treatment to determine whether it influences the dependent measure[s], followed by withdrawal of treatment to determine whether performance returns to or near baseline levels, followed by reintroduction of treatment to determine whether performance is again influenced).

A second manipulation (multiple baselines) involves the staggered introduction of the treatment. This can involve both within- and between-subjects comparisons. For example, after researchers establish a stable baseline pattern of performance for each participating student, treatment is implemented with one student to determine whether it influences the students’ performance in a predictable fashion. If the instructed student demonstrates the predicted pattern of change, while uninstructed students’ performance remains unchanged from baseline, then the treatment is implemented with the next student to determine whether the pattern described above is replicated. This systematic delay in introduction of the treatment continues until all students receive instruction.

With both of these manipulations, experimental control is established only if performance on the dependent measures is stable during each experimental phase (e.g., baseline and treatment) and there is no trend in the pattern of baseline performance in the direction predicted by the intervention. In addition, it is generally agreed that experimental control is not established until there are at least three demonstrations that the manipulation had the predicted impact. Thus, experimental control is established when the predicted covariation between the introduction of the treatment and changes in the dependent measure(s) are demonstrated through at least three demonstrations in an experiment.

In contrast to true- and quasi-experimental designs where the focus is on group performance, single subject design examines the

effectiveness of a treatment at the individual level (single subject design studies, however, typically include more than a single student). Considerable emphasis is placed on fully describing the participants, the context in which the investigation took place, and factors that influence participants’ performance prior to intervention (Horner et al., 2005). Such rich descriptions set the stage for determining the boundaries of generalization, but a basic tenet in single subject design research is that the generality of a treatment is not established in a single study. Instead, external validity is established by systematically replicating effects across multiple participants, locations, and researchers.

Why a Meta-Analysis of Single Subject Design Studies Is Important

There are three reasons why it is advantageous to conduct a systematic, empirical, and full review of single subject design writing intervention research. First, the meta-analyses of true- and quasi-experimental investigations of writing interventions have identified only 12 interventions that improve the writing of elementary and secondary students (see Bangert-Drowns, 1993; Bangert-Drowns et al., 2004; Goldring et al., 2003; Graham, 2006a; Graham & Harris, 2003; Graham & Olinghouse, in press; Graham & Perin, 2007a, 2007b; Hillocks, 1986). These include (ranked according to the magnitude of their impact): teaching strategies for planning, revising, and editing; teaching written summarization; having students help each other carry out specific writing processes such as planning, drafting, or revising/editing; establishing specific goals for students’ writing; teaching handwriting, spelling, or typing to students; using word processing as a medium for writing; teaching students how to write more complex sentences through sentence combining activities; engaging students in inquiry as a means for developing ideas for writing; encouraging students to engage in prewriting activities to gather and organize possible writing content; establishing a process approach to writing; using writing as a tool for content learning; and having students study and emulate models of good writing. While the identification of these 12 treatments is an important accomplishment, these practices do not cover all aspects of learning to write. A meta-analysis of single subject design writing interventions has the potential to broaden current evidence-based recommendations.

Second, a meta-analysis of single subject design studies also has the potential to strengthen, undermine, or nuance the trust that can be placed in one or more of the 12 writing treatments identified as effective above. For example, converging evidence from true- and quasi-experimental studies as well as single subject design research would bolster the claim that a treatment was or was not effective, whereas conflicting evidence would undermine such a claim or foster a more nuanced conclusion. The need for additional evidence is also supported by the fact that only 4 of the 12 treatments identified above (teaching writing strategies, process writing approach, word processing, and using writing to learn) were based on 10 or more studies.

Third, most of the true- and quasi-experimental writing intervention research has been conducted with students representing the full range of writing ability in a typical classroom (Graham & Perin, 2007a). The only exceptions to this involve strategy instruction, word processing, and setting goals for students’ writing. In

contrast, single subject design studies often involve students' experiencing difficulty. Consequently, a meta-analysis of single subject writing intervention research has the potential to increase the number of identified evidence-based practices for struggling writers.

It is also important to establish that there is currently no comprehensive meta-analysis of single subject writing intervention studies. Graham and colleagues (Graham, 2006a; Graham & Harris, 2003) conducted meta-analyses of single subject design studies examining strategy instruction in planning, revising, and/or editing. The first of these reviews (Graham & Harris, 2003) only focused on a specific model of strategy instruction, whereas the second review (Graham, 2006a) concentrated on all single subject design studies in this area. In both reviews, teaching writing strategies had a positive impact on writing, providing additional support to the findings from true- and quasi-experimental studies (e.g., Graham & Perin, 2007a).

The Current Meta-Analysis

The meta-analysis reported in this article draws on but greatly extends the two previous meta-analyses of single subject strategy instruction research in writing (Graham, 2006a; Graham & Harris, 2003). We not only broadened the review to include all writing practices tested via single subject design but conducted a broader search than Graham (2006a) or Graham and Harris (2003), resulting in the identification of some studies not included in their review. To reduce the likelihood of only identifying studies where positive effects were obtained (i.e., nonsignificant findings are rarely published in peer-refereed journals), we searched as broadly as possible, including studies published in journals as well as dissertations, theses, and book chapters.

The primary research question guiding this review was, Which writing practices tested via single subject design procedures are effective with students in Grades 1–12? Consistent with Graham and Perin (2007a), no conclusions were drawn about a treatment unless there were at least four studies with a conceptually similar outcome measure assessing its impact. Furthermore, we assessed the quality of each single subject design study included in this review, using quality indicators developed by Horner et al. (2005). This allowed us to identify strengths and weaknesses in the current body of single subject writing intervention research and temper conclusions about the effectiveness of a treatment based on the quality of the research.

The theoretical bases for many of the writing treatments in this review were grounded in behavioral theory. This is not surprising, as single subject design methodology grew out of behavioral work in the 1960s (Horner et al., 2005). Nevertheless, some treatments, such as strategy instruction, were influenced by cognitive (Hayes, 2000) and/or behavioral models, whereas the theoretical underpinnings of other treatments were unstated (e.g., word processing). Consequently, we draw no claims about the validity of specific theories.

Method

Location and Selection of Studies

The strategies that we used to locate and select studies for this meta-analysis were influenced by five factors. First, studies were

included that involved students in Grades 1–12. This differed from Hillocks's (1986) comprehensive meta-analyses of true- and quasi-experimental writing intervention research that focused on students in Grade 3 to college, and Graham and Perin (2007a, 2007b), who concentrated on Grades 4–12. As noted earlier, our goal was to identify effective writing treatments for both elementary and secondary students.

Second, we included studies that were conducted with students attending regular public schools, private schools, alternative schools, summer programs, clinics, and residential centers. We cast a broader net than Graham and Perin (2007a, 2007b), who did not include studies conducted in special schools for persons with disabilities (e.g., residential centers), as we were especially interested in identifying writing practices that would be effective with struggling writers, including those with disabilities (we did not purposefully exclude any type of disability).

Third, we included only single subject designs with mechanisms for establishing experimental control (i.e., demonstrating a functional relationship between the independent and dependent variable). This included reversal and multiple baseline designs (described in the introduction) as well as alternating treatment and changing criterion designs. Alternating treatment designs involve the rapid alternation of two or more distinct treatments (with each treatment presented the same number of times) in a counterbalanced fashion to determine their impact on a single outcome measure. The treatments are alternated rapidly to reduce the possibility of carryover effects, and treatments are counterbalanced to eliminate order effects. We only included alternating treatment design studies that included a baseline (this is a recommended practice but is not absolutely essential). Baseline is established before introduction of the alternating treatments or by including it as one of the alternating treatment conditions (i.e., a no-treatment condition). Experimental control is established when a treatment repeatedly produces the predicted change in behavior in reference to baseline (whether this is a baseline established before alternating treatments or a no-treatment control included as one of the alternating treatments).

With a changing criterion design, the desired outcome during treatment is gradually and systematically increased or decreased for a specific behavior. After establishing a stable baseline pattern of performance, the researcher introduces the treatment (e.g., reinforcement) but divided into subphases, with two or more subphases requiring increased changes (if the desire is to improve the behavior; e.g., "Write 20 words," "Write 30 words") and at least one subphase requiring decreased changes (e.g., "Now write only 15 words"). Such changes are meant to gradually move the student toward the desired terminal goal (e.g., "Write 50 words") while demonstrating that it is treatment, and not maturation, that causes the change. Experimental control is established if the student's behavior matches or exceeds the predetermined criterion specified in each subphase of treatment.

Fourth, studies were included if they provided the data needed to calculate the effect size, percentage of nonoverlapping data (PND; this measure is described more fully later). This required that the data at each assessment point was provided either in graph or tabular form. If a study did not have such data for at least one writing measure, it was not included in this review. In addition, we did not include studies where baseline data were not collected. Nor did we include studies that focused solely on the teaching of

handwriting or spelling, as they typically do not examine whether such instruction impacts students' actual writing (Graham, 1999).

Fifth, we searched broadly to identify possible studies. This involved locating peer-refereed and nonrefereed studies from a variety of sources, including studies in prior meta-analyses (i.e., Graham, 2006a; Graham & Harris, 2003), journals, theses and dissertations, conference proceedings, and books. We searched broadly, going beyond published peer-refereed articles, to reduce the possibility of bias, as studies are rarely published in peer-refereed journals when they do not obtain positive treatment results.

A number of databases were searched during January 2007 for relevant studies, including ERIC, PsycINFO, ProQuest, Education Abstracts (i.e., Education Full Text), and Dissertation Abstracts. We ran multiple searches in these databases, pairing writing, composition, and narrative with the following terms: *dictation, genre, genre and instruction, goal setting, grammar, inquiry, mechanics, models, peer collaboration, peer planning, peer revising, peers, planning, revising, pre-writing activities, process writing, reinforcement, self-evaluation, self-monitoring, sentence combining, sentence construction, speech synthesis, spell checkers, strategy instruction, summary instruction, summary strategies, summary writing, technology, usage and mechanics, word processing, word processor, writer's workshop, and writing assessment*. These terms were also included in the search by Graham and Perin (2007a).

These same databases were searched to identify single subject design writing intervention research conducted by 11 prominent researchers in this area (e.g., K. Ballard, T. Glynn, and C. MacArthur). We also conducted a hand search of the following journals that frequently publish single subject design research: *Education and Treatment of Children, Educational Psychologist, Exceptional Children, Journal of Applied Behavior Analysis, Journal of Behavioral Education, Journal of Learning Disabilities, Journal of Special Education Technology, Learning Disabilities & Practice, Learning Disability Quarterly, and Remedial & Special Education*. Finally, the reference lists for all obtained articles were searched to identify relevant studies.

Using these search mechanisms, we collected 119 documents and found 88 studies that were suitable for this review. The most common reason for excluding a study was that it did not include a writing outcome measure. This occurred in 11 instances. Nine studies were eliminated because they did not apply a design where experimental control could be established, whereas 5 studies were excluded because they contained no baseline phase. Finally, 3 studies were omitted because the same data were published elsewhere, 2 did not provide the data needed to calculate PND, and 1 study included a combined writing/reading intervention.

Categorizing Studies Into Treatment Conditions

First, each study was read and placed into a treatment identified in advance. These treatments corresponded to the descriptors used in the electronic searches. These categories were supplemented by two additional treatments common in behavioral research: reinforcement and direct instruction. Studies that did not fit neatly into one of these preidentified categories were held apart until all studies were read and sorted. At this point, the studies in each preidentified treatment were reread to determine whether the in-

tervention in each study represented the same general treatment. If this was not the case, they were placed with the studies that were not classified during the initial reading. All of the studies in the unclassified pile were read again, resulting in the construction of new treatments. The studies in these new treatments and any preidentified treatment where an additional study was placed were again reread to determine whether each intervention represented the same general treatment. As this process took place, we refined some initial treatment categories and eliminated others (when no studies tested that treatment).

At least one or more studies examined the effectiveness of the following 20 treatments (ordered from most to least frequently studied): strategy instruction for planning/drafting (25 studies), self-monitoring (8 studies), goal setting for productivity (7 studies), reinforcement (6 studies), prewriting activities (5 studies), sentence construction (5 studies), strategy instruction editing (5 studies), strategy instruction paragraph construction (5 studies), word processing (5 studies), teaching grammar/usage (4 studies), feedback on writing (4 studies), strategy instruction other (4 studies), strategy instruction revising (2 studies), direct instruction of a broad array of skills (2 studies), goal setting for grammar/sentence construction (2 studies), word processing plus (2 studies); these studies included word processing plus an additional support, for example, text read aloud by the computer), dialogue journals (1 study), direct teaching of self-regulation strategies (1 study), repeated writing (1 study), and verbal encouragement (1 study).

The most common treatment involved teaching strategies for carrying out specific aspects of the writing process. Instead of lumping these into a single category, we separated them into different categories depending on the processes emphasized. Most of these studies focused on teaching a genre-specific writing strategy for planning and drafting papers (stories, persuasive essays, and expository text), and this formed a single category. All of the planning/drafting strategies in these 25 investigations were taught using the Self-Regulated Strategy Development model (SRSD; see Harris & Graham, 1996, for a description of this teaching approach).

The other strategy instructional categories that we created concentrated on teaching strategies for writing paragraphs, editing only, or revising/editing. Since these investigations emphasized different processes (from each other and the strategy for planning/drafting studies), we created a category for each. Finally, a small number of strategy instructional studies did not fit neatly into any of these categories and were placed in a strategy instruction other category. Studies in this category included investigations where strategy instruction was combined with other practices (such as a token reinforcement system; Boyer, 1990), students were taught a summary writing strategy (Nelson, Smith, & Dodd, 1992), or the study did not assess students' independent use of the strategy (Li, 2000). Our decision to create multiple strategy treatments differed from Graham and Perin (2007a, 2007b), where all strategy studies were included in the same category.

We also differed from Graham and Perin (2007a, 2007b) in that we created two treatment categories for goal setting. The purpose of the goals in these two categories differed, as one set of goals was aimed at increasing productivity, and the other focused on grammatical correctness. It should be noted that we created two word processing treatment categories. This included word processing, which examined the effects of word processing and software

commonly bundled into such a program (e.g., spell checkers) as well as word processing plus, which included word processing plus programs like word prediction (Handley-More, Deitz, Billingsley, & Coggins, 2003) and speech synthesis programs that read written text aloud (Channon, 2004).

As noted earlier, we calculated only a summary statistic (mean PND) for treatments that included four or more studies with a conceptually similar outcome measure. Table 1 presents a definition for the 10 treatment categories that met this condition. We decided to adopt this criterion, as it was used by Hillocks (1986) in his seminal review and Graham and Perin (2007a, 2007b) in their more recent review. If reviewers use similar criterion, it is easier to make valid comparisons across reviews. We recognize, however, that small sample sizes are less reliable and must be interpreted more cautiously than a summary statistic based on a larger number of studies.

The decision to stress a conceptually similar outcome measure in computing a summary statistic also has precedence in the two prior comprehensive reviews of true- and quasi-experimental research (Graham & Perin, 2007a, 2007b; Hillocks, 1986). In those reviews, the focus was only on the outcome measure of writing quality (it was assumed that there would be less noise or error in the analyses, if the outcome measures were conceptually similar). With the single subject design studies reported here, researchers typically graphed just the measure that was most directly associated with the intervention (e.g., if the intervention was meant to increase writing output, then this variable was graphed). Thus, there was no single outcome measure for most of the studies in this review. For half of the treatments described in Table 1, there were at least four or more studies that graphed writing productivity (number of words and sentences). These treatments were goal setting for productivity, reinforcement, self-monitoring, strategy instruction for planning/drafting, and word processing. Two treatments (prewriting and strategy instruction for planning/drafting) met the four study criteria for quality of writing (as measured with holistic rating scales; see Graham & Perin, 2007a), whereas an-

other two treatments (strategy instruction for planning/drafting and strategy instruction for paragraph construction) met it for elements (basic structural elements of a genre or writing task, such as story parts). Sentence construction, strategy instruction for editing, and teaching grammar/usage had four or more studies that graphed number of complete sentences, errors corrected, and grammar errors, respectively.

Coding of Study Features

Descriptive information. For each study that met the established criteria, the following 10 pieces of information were collected and coded: type of design (i.e., withdrawal design, multiple-baseline design, alternating treatment design, and changing criteria design), number of participants, type of writers (i.e., full range of writers in a typical classroom, above average writers only, average writers only, struggling writers only, and English language learners), grade of study participants, age of study participants, disability status of each participant (i.e., attention-deficit/hyperactivity disorder [ADHD], behavior disorder, emotional and behavior disorder, emotionally disturbed, learning disability, mild language delay, mild mental retardation, other health impairment, orthopedic impairment, speech and language delay, visual impairment, and Section 504), race/ethnicity of participants (i.e., African American, American Indian, Caucasian, Hispanic, Arab, or Other), geographical location of study (i.e., suburban, rural, or urban), person providing instruction (i.e., teacher or member of research team), and written description of independent variable (coded as one of the 20 treatment categories established in the previous section).

Quality indicators. In addition to this descriptive information, each study was evaluated to see whether it met 11 specific indicators of study quality proposed by Horner et al. (2005) for single subject design research. Some of these quality indicators were based in part on the descriptive information collected and coded above, but others required the collection of additional information.

Table 1
Summary of Evidence-Based Practices for Teaching Writing

Treatment category	Definition
Strategy instruction: Planning/drafting	Students were taught strategies for planning, drafting, and/or revising stories, persuasive essays, and/or expository essays. This included modeling of the strategy and guided practice to facilitate independent use of it.
Teaching grammar/usage	Students were taught grammar and usage skills via well-sequenced and highly focused instruction.
Goal setting for productivity	Students were provided or set an explicit goal to write more.
Strategy instruction: Editing	Students were taught a strategy for editing their papers. This included modeling of the strategy and guided practice to facilitate independent use of it.
Word processing	Students used word processing as their medium for writing.
Reinforcement	Students received praise, public recognition, or tangible items contingent on writing improvement(s).
Prewriting activities	Students used graphic organizers, including story maps and outlines, for generating ideas prior to writing.
Sentence construction	Students were taught sentence skills through either sentence combining or a strategy for writing sentences.
Strategy instruction: Paragraph construction	Students were taught a strategy for writing a paragraph. This included modeling of the strategy and guided practice to facilitate independent use of it.
Self-monitoring	Students self-monitored either their on-task behavior, writing productivity, or writing quality, and the results of their performance was displayed.

The score for each indicator ranged from 0 to 1 (with a score of 1 indicating that the quality indicator was met), and all scores for a study were summed to obtain a total quality score.

The quality indicator, participant description, was based on five or six pieces of information, depending on whether students with disabilities were included in the study. A study received a score of 1 on this indicator, if information was provided on participants' (a) ages or grades, (b) gender, (c) socioeconomic status, (d) ethnicity, (e) writing achievement at the start of the study, and when appropriate (f) disability status. The score for this indicator was calculated by assigning 1 point for each of these items if they were reported, and then dividing by either 5 or 6, depending on whether students with disabilities were included in the study.

A score of 1 was awarded if the researcher described the procedures for determining how students were selected to participate in the study (this needed to be done in enough detail so that another researcher could replicate these procedures). A score of 1 was also assigned if the location where instruction took place was described in enough detail so that it could easily be visualized. For each indicator, a score of 0 was awarded if the criterion was not met.

For the quality indicator, operationally defining dependent variables, we examined whether the procedures for scoring dependent variables (that were graphed) were described in enough detail to allow other researchers to use these measures. If all dependent measures were defined in this fashion, a score of 1 was assigned. A score of 0 was assigned if no dependent measures met this criterion. A score of .50 was awarded if at least one of the dependent variables met this criterion.

The quality indicator, interrater reliability of graphed dependent variables, was assessed by calculating the proportion of dependent measures that were scored reliably. If reliability for all dependent measures equaled or exceeded .60, a score of 1 was awarded. A score of 0 was awarded if none of the dependent variables met this criterion. If the reliability for some but not all dependent variables was .60 or higher, then the number of dependent measures meeting criterion was divided by total number of dependent variables. Researchers used different means to calculate reliability (ranging from percentage agreement to correlation coefficients).

For the quality indicator, multiple baseline data points, we examined each baseline to determine whether it had three or more data points. A score of 1 was assigned if all baselines met this criterion, whereas a score of 0 was assigned if no baselines met it. If some but not all baselines met this criterion, then the proportion that met it was calculated. The exact same procedures were used to score the quality indicator, multiple intervention data points.

The remaining four quality indicators (treatment description, fidelity of treatment, testing procedure descriptions, and social validity) received a score of either 1 or 0. A score of 1 was awarded for each of the following: a treatment was described in enough detail so that it could be applied by others, the researcher collected and reported data that demonstrated fidelity of treatment (treatment was delivered as intended), enough detail (e.g., directions) was provided so that other researchers would be able to administer the writing assessments, and data on the social validity of the treatment were collected (as suggested by Kennedy, 2002, this could include questioning students, teachers, or others about the social validity of the treatment or a demonstration that treat-

ment gains were maintained for 3 or more weeks following instruction).

In addition to these 11 quality indicators, we also examined whether experimental control was established. This is critical to establishing a functional relation between the independent and dependent variables. Experimental control was established if there were at least three demonstrations in a study of an experimental effect that met these criteria: (a) the last three data points in baseline established a stable pattern of behavior (i.e., the last three data points fluctuated by no more than 20% from the highest score of the three data points), (b) there was no trend in baseline performance in the direction predicted by the treatment, (c) and treatment had the predicted impact on behavior. To determine whether treatment had the predicted impact on behavior (letter *c* above), 50% or more of the treatment data points had to exceed the strongest baseline score.

In determining whether experimental control was established, it was necessary to customize the third criterion above to the four different single subject designs included in this review. For withdrawal designs, treatment had the predicted impact on behavior when 50% or more of the data points in the first introduction of the treatment (B_1) surpassed the strongest initial baseline score (A_1) and strongest score when treatment was withdrawn and the second baseline was initiated (B_2). For multiple baseline designs, 50% or more of the treatment data points had to exceed the strongest baseline score, and there could be no corresponding improvement or trend toward improvement for each baseline that had not received treatment yet. For alternating treatment designs, 50% or more of the treatment scores for a specific treatment had to exceed the strongest score from baseline. This included the strongest score from either a baseline established before treatments were alternated and/or baseline scores collected as an alternating condition (i.e., no treatment). For changing criterion designs, all of the treatment data points from the first subphase of treatment had to fall within 20% of the stated goal for that subphase, with this same criterion applied to subsequent subphases.

Reliability. To establish the reliability of scoring procedures, 20% of the studies ($n = 18$) were randomly selected and rescored. The second reader was trained and scored all 18 studies independent of Leslie Rogers. Interrater reliability was established for each scoring procedure separately. The scoring procedures were reliable, as percentage of agreement ranged from 89% to 100%, with a mean percentage of agreement across all variables of 96%.

Calculation of Effect Sizes

In order to assess overall treatment effects, we employed a nonparametric approach to meta-analysis and calculated the percentage of nonoverlapping data points between baseline and treatment phases: referred to as PND (Scruggs, Mastropieri, & Casto, 1987). PND is the percentage of data points in treatment that represent an improvement over the most positive value obtained during baseline. As recommended by Scruggs, Mastropieri, Cook, and Escobar (1986), PND was not calculated when ceiling or floor levels were evident during baseline (depending on the intended direction of the outcome variable), as there was no room for treatment effects to be realized.

For each study, we calculated a separate PND for each measure that provided the needed information. PND for a specific measure

in a study was calculated by first obtaining PND for each baseline treatment comparison (or changing condition) and then calculating an average PND across all relevant changing conditions for that measure. When possible, we calculated a PND for treatment, maintenance, and generalization for each measure, as we were interested not just in immediate effects but in the impact of the treatment over time as well as transfer effects to other situations. A PND calculated for treatment involved comparing scores during treatment, immediately following treatment, or both to the strongest baseline score for each changing condition in the study. A PND calculated for maintenance involved comparing a student's scores collected 3 weeks or more after treatment ended to the strongest baseline score for each changing condition in the study. A PND for generalization (typically to another genre or setting) involved comparing a student's generalization scores during or after treatment to the strongest generalization score during baseline for each changing condition in the study.

When there were four or more studies of a treatment that included a conceptually similar outcome measure, we calculated a mean, median, and range of PNDs for that measure across studies (confidence intervals were not calculated as PND lacks a known sampling distribution; Parker, Hagan-Burke, & Vannest, 2007). When possible, we calculated these three indices (mean, median, and range) for treatment, maintenance, and generalization. PND was interpreted using criteria proposed by Scruggs et al. (1986): PND greater than 90% is a large effect, PND between 70.1% and 90% is a moderate effect, PND between 50.1% and 70% is a low or small effect, and PND 50% or below is classified as not effective.

To avoid overinflating the importance of a single study, it is recommended that a single effect size be calculated for each study (Lipsey & Wilson, 2001). Although we computed more than one PND for many studies, this basic concept was followed here. For all but one of the treatments, only one effect size from each study was used to calculate summary PNDs. In all of these instances, we were only able to calculate a summary PND for a single measure for each treatment. The only exception involved the treatment of strategy instruction for planning/drafting. Almost all of the studies in this category had a conceptually similar outcome measure that was graphed (i.e., elements), but there were also enough studies that conjointly graphed productivity and writing quality that we were able to calculate a separate summary PND for these measures too. We decided to present a summary PND for these measures separately, as it made little sense to develop a single summary PND combining all three measures or simply ignore two of them. The summary PND for a measure was never based on more than one effect size from a single study, however. The same was true when a summary PND was computed for treatment, maintenance, or generalization.

PND was calculated for all studies by Leslie Rogers, and reliability was established by a second rater who calculated PND for 18 randomly selected studies. Reliability was .99.

Why did we decide to use PND over other methods for calculating effect sizes for single subject design studies? First, as Scruggs and Mastropieri (1998) noted, single subject design effect sizes created by subtracting mean baseline performance from mean treatment performance and dividing by baseline (or the pooled) standard deviation (much like Cohen's *d*) do not take into account the within-subject nature of the data and can result in effect sizes

that are idiosyncratic and meaningless (e.g., when we have used this method for SRSD single subject design studies in writing, the effect sizes are typically 3.0 and higher; they average just over 1.0 in true- and quasi-experimental studies; Graham & Perin, 2007a). Other alternative methods, such as the regression effect size method (which examines the proportion of student score variance explained by phase differences), also possess limitations that made them less suitable for our analysis (see Scruggs & Mastropieri, 1998, for a discussion of different approaches). For example, the parametric assumptions (normality, equal variance, and serial independence of data) underlying techniques such as the regression effect method are not commonly met in single subject design studies (PND is not bound by these parametric assumptions; see Parker et al., 2007). Second, we would have had to eliminate many studies from this review if the regression approach were applied, as it was not possible to determine the exact scores for each data point from graphs in many studies (it was possible, however, to compute PND, as it is only necessary to establish that one score was greater than another). Third, PND was the most commonly used measure in previous meta-analyses of single subject writing intervention research; Graham, 2006a; Graham & Harris, 2003), making it easier to compare findings across reviews.

Results

Table 2 presents basic information about the individual studies that tested the effectiveness of the 10 treatments that included four or more studies with a conceptually similar outcome measure (see Table 1 for a description of these treatments). This includes information about study design, participants (number, type of writer, grade, age, and race/ethnicity), geographic location, overall quality of study (based on the 11 quality indicators), whether experimental control was established, as well as the PNDs for specific measures. The Appendix includes information about studies that were included in treatments with three or less studies with a conceptually similar outcome measure. Table 3 presents summary PNDs (mean, median, and range) for each of the treatment categories included in Table 2, whereas Table 4 presents summary PNDs for the treatment, strategy instruction for planning/drafting, by type of genre (story versus expository).

Table 5 presents the average quality score for studies in each of the 10 treatments included in Table 2. The total quality score was the sum of the scores for the 11 quality indicators (e.g., participant description and participant selection). For each treatment, we also report the percentage of studies where each quality indicator was met. This same information is presented cumulatively for all studies included in the 10 treatments, providing a general indication of the quality of the single subject design intervention studies on which our findings are based.

Before examining individual treatments, we first offer some comments on the overall quality of single subject design writing intervention studies. As can be seen in Table 5, almost all studies quantified dependent variables, established reliability of variables, collected three or more data points in baseline as well as intervention, and adequately described the intervention. However, participant description and selection were only adequately described in about one half of the studies, and an adequate description of the

(text continues on page 892)

Table 2
Writing Instruction Treatments That Included Four or More Studies

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Instr.	Quality score (0–11)	PND TX	PND POST	PND MAINT	PND GEN
Stories Adkins (2005)	MBD	3	SW	2 to 3 (7&9)	EBD	AA	SUB	RES	10.00	100% ^a (Elements)	100% ^a (Elements)	100% (Elements)	100% (Elements)
Albertson: Study 1 (1998)	MBD	4	SW	5 to 8 (10–13)	ADHD (n = 1)	NR	NR	RES	6.67	95% (Elements)	68% ^a (Production)	100% (Elements)	100% (Elements)
Albertson: Study 2 (1998)	MBD	2	SW	6 (11–12)	—	NR	NR	Teacher	6.60	100% (Grammar)	100% (Elements)	67% (Quality)	100% (Elements)
Albertson & Billingsley (1997)	MBD	2	AVG	6 (12)	Gifted	NR	NR	RES	7.08	67% (Elements)	100% (Grammar)	100% (Elements)	100% (Elements)
Danoff et al. (1993)	MBD	6	FR	4 to 5	LD (n = 3)	Asian (n = 1) H (n = 1) C (n = 4)	SUB	Teacher	8.67	100% (Production)	100% ^a (Elements)	100% (Elements)	100% (Elements)
Lane et al. (2007)	MBD	6	SW	2 (7–8)	—	AA (n = 2) C (n = 4)	RURAL	RES	10.00	100% ^a (Elements)	100% ^a (Elements)	100% (Elements)	67% (Elements)
Lienemann et al. (2006)	MBD	6	SW	2 (7–8)	ADHD (n = 1) LD (n = 1) OI (n = 1)	AA (n = 1) C (n = 4) H (n = 1)	RURAL	RES	10.00	97% ^a (Elements)	83% (Elements)	83% (Elements)	67% (Elements)
Reid & Lienemann (2006)	MBD	3	SW	3 to 4 (9–10)	ADHD (n = 1)	C	RURAL	RES	11.00	100% (Elements)	100% (Elements)	100% (Elements)	100% (Elements)
Saddler et al. (2004)	MBD	6	SW	2 (7–8)	MLD (n = 1) SLI (n = 1)	AA	SUB	RES	10.00	100% ^a (Elements)	100% ^a (Elements)	88% (Elements)	82% (Elements)
Germain (2004)	MBD	10	AVG	4 (9–10)	—	AA (n = 2) H (n = 1) C (n = 7)	SUB	RES	7.00	95% ^a (Quality)	100% ^a (Quality)	100% (Quality)	100% (Quality)

Table 2 (continued)

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Instr.	Quality score (0–11)	PND TX	PND POST	PND MAINT	PND GEN
Saddler (2006)	MBD	6	SW	2 (7)	LD	AA	URBAN	RES	9.00	100% ^a (Elements)	100% ^a (Elements)	100% (Elements)	
Saddler & Asaro (2007)	MBD	6	SW	2 (7)	LD	AA (n = 4) C (n = 2)	URBAN	RES	8.00	100% ^a (Production)	100% ^a (Quality)	83% (Production)	
Harris & Graham (1985)	MBD	2	SW	— (12)	LD	NR	SUB	RES	8.67	96% ^a (Production)	33% (Production)	100% (Production)	
Korducki (2001)	MBD	8	ESOL	5 (10–11)	LD (n = 4)	H	URBAN	RES	9.83	92% (Grammar)	92% (Grammar)	75% (Grammar)	89% (Grammar)
Guzel-Ozmen (2006)	MBD	4	SW	— (13–17)	MMR	Arab	NR	RES	7.83	100% ^a (Structure Element)	100% ^a (Structure Element)	100% (Structure Element)	Less Dominant Language
Troia et al. (1999)	MBD	3	SW	5 (10–11)	LD	AA (n = 1) C (n = 2)	SUB	RES	9.66	100% ^a (Elements)	100% ^a (Elements)	67% (Elements)	67% (Elements)
Persuasive essays													
De La Paz & Graham (1996)	MBD	3	SW	5	LD	AA (n = 2) C (n = 1)	SUB	RES	8.83	100% ^a (Elements)	100% ^a (Elements)	100% (Elements)	
Graham & Harris (1989)	MBD	3	SW	6 (12)	LD	NR	SUB	RES	8.67	100% ^a (Elements)	100% ^a (Elements)	89% (Elements)	100% (Elements)
Graham et al. (1992)	MBD	4	SW	5 (11–12)	LD	NR	URBAN	RES	8.67	100% ^a (Elements)	100% ^a (Elements)	100% (Elements)	Setting 89% (Elements)
Lienemann (2006)	MBD	4	SW	4 to 5 (9–11)	LD (n = 1)	C	RURAL	RES	10.00	100% ^a (Elements)	100% ^a (Production)	100% (Elements)	100% (Elements)
Mason & Shriner (2007)	MBD	6	SW	2 to 5 (8–12)	EBD (n = 4) LD (n = 1) 504 (n = 1)	NR	NR	RES	8.75	89% ^a (Elements)	89% ^a (Elements)	100% (Elements)	100% (Elements)
Sexton et al. (1998)	MBD	6	SW	5 to 6 (10–12)	LD (n = 1)	AA (n = 5) C (n = 1)	SUB	RES	8.00	68% ^a (Elements)	68% ^a (Elements)	25% (Elements)	100% (Elements)
Expository essays													
De La Paz (2001)	MBD	3	SW	7 to 8	ADD (n = 1) SLI (n = 2)	C	SUB	Teacher	9.00	100% ^a (Elements)	100% ^a (Elements)	100% (Elements)	Setting

(table continues)

Table 2 (continued)

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Instr.	Quality score (0-11)	PND TX	PND POST	PND MAINT	PND GEN
De La Paz (1999)	MBD	22	FR	7 to 8	LD (n = 6)	AA (n = 2) C (n = 20)	SUB	Teacher	10.00		95% ^a (Elements) 78% ^a (Production) 97% ^a (Quality)	100% (Elements) 88% (Production) 100% (Quality)	
Mason et al. (2006)	MBD	9	SW	4 (9-10)	EBD (n = 1) LD SLI (n = 2)	AA (n = 1) C (n = 8)	URBAN	RES	9.83		94% ^a (Production)	100% (Production)	84% (Production) Reading
Teaching grammar/usage													
Campbell et al. (1991)	ABAB & MBD	3	SW	2 (9)	MMR (n = 1) LD (n = 2)	AA (n = 1) C (n = 2)	NR	Peers	9.00	79% (Grammar)		100% (Grammar)	77% (Grammar) Other Untrained Peers
Dowis (1991)	MBD	14	SW	5 to 6	LD	NR	SUB	Teacher	8.50	75% ^a (Grammar)	87% (Grammar)	89% (Grammar)	
Dowis & Schloss (1992)	MBD	4	SW	6 (12)	LD	AA (n = 1) C (n = 3)	SUB	Teacher	8.00	88% ^a (Grammar)		76% (Grammar)	
Herman et al. (1976)	MBD	9	SW	6	—	NR	NR	Teacher	7.20	88% ^a (Grammar)	83% (Grammar)		
Goal setting for productivity													
Hopman & Glynn (1989)	CCD	4	SW	HS (13)	—	NR	URBAN	RES	6.40	26% (Production)	63% (Production)		
Kastelen et al. (1984)	MBD	16	FR	8 (13)	—	NR	NR	Teacher	6.20	100% (Production)		100% (Production)	
Seabaugh & Schumaker (1994)	MBD	3	FR	9 to 12 (14-18)	LD (n = 2)	NR	NR	RES	10.83	60% (Production)			
Van Houten et al. (1974)	ABAB	55	FR	2 & 5	—	NR	NR	Teacher	6.40	94% ^a (Production)			
Van Houten & McKillop (1977)	ABAB	38	FR	10 & 11	—	NR	NR	Teacher	6.40	91% ^a (Production)			
Weygant (1981)	MBD	9	SW	4 to 5 (10-12)	LD	C	NR	RES	9.00	99% (Production)			
Wolfe (1997)	CCD	4	SW	2 to 3 (9)	LD	NR	URBAN	Teacher	9.50	63% ^a (Production)			
Strategy instruction (editing)													
Beals (1983)	MBD	9	FR	10 (15-16)	LD (n = 3)	AA (n = 4) C (n = 5)	URBAN	Teacher	7.75		71% (Errors Corrected)		
McNaughton et al. (1997)	MBD	3	SW	10 & 12 (16-18)	LD	NR	RURAL	RES	7.67	95% ^a (Errors Corrected)	100% (Errors Corrected)		

Table 2 (continued)

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Instr.	Quality score (0-11)	PND TX	PND POST	PND MAINT	PND GEN
Schmidt (1983)	MBD	7	SW	10 to 12 (14-16)	LD	C	SUB	RES	7.00		50% (Errors Corrected) 29% (Elements)	34% (Errors Corrected) 34% (Elements)	
Schumaker et al. (1982)	MBD	9	SW	8 to 12 (12-18)	LD	NR	NR	Teacher	6.34		100% (Errors Corrected) 100% (Errors Detected) 100% (Errors / Word)		
Scott (1993)	MBD	4	SW	4 (10-12)	LD	AA (n = 3) C (n = 1)	URBAN	RES	10.92		100% ^a (Errors Corrected)		
Brigman (1994)	ATD	6	SW	1 (6-7)	—	NR	Word processing SUB & RURAL	Teacher	9.90	63% ^a (Production)			
Christensen (1993)	MBD	12	SW	4	Gifted	C	URBAN	Teacher	5.00	100% (Production)			
Handley-More et al. (2003)	ATD	3	SW	4 to 5 (10-11)	LD	NR	NR	RES	5.50	29% (Production)			
Jones (1985)	ABAB	6	SW	4 (9)	—	AA (n = 3) C (n = 3)	RURAL	RES	7.60	86% ^a (Production) 62% ^a (Holistic Content) 14% (Holistic Form)			
Burnett (1984)	ABAB	10	SW	1-3 & 5	—	C	RURAL	Teacher	8.80	64% ^a (Quality)			
Ballard & Glynn (1975)	MBD	14	FR	3 (8-9)	—	NR	RURAL	Teacher	8.20	100% ^a (Production)			
Blasé-Maloney & Hopkins (1973)	MBD	14	NR	4 to 6	—	NR	NR	Teacher	6.20	100% ^a (Production)			
Blasé-Maloney et al. (1975)	MBD	19	NR	3	—	NR	URBAN	Teacher	6.20	100% (Production)			
Brigham et al. (1972)	MBD	13	SW	5	—	NR	NR	Teacher	6.40	84% (Production)			
Bording et al. (1984)	MBD	9	SW	(12-16)	BD (n = 5) MMR (n = 2) LD (n = 2)	C	NR	Teacher	8.83	57% (Grammar)			
Newstrom et al. (1999)	MBD	1	SW	9	EBD	NR	NR	NR	5.50	100% (Grammar)			
Campbell & Willis (1978)	MBD	32	FR	5 (10-12)	—	NR	URBAN	Teacher	7.60	86% ^a (Creativity)			

(table continues)

Table 2 (continued)

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Instr.	Quality score (0-11)	PND TX	PND POST	PND MAINT	PND GEN
Channon (2004) Inspiration Channon (2004) Draft: Builder Thanhouser (1994)	ATD	7	SW	8 (14-15)	LD	H	URBAN	Teacher	8.83	84% ^a (Quality)			
	ATD	7	SW	8 (14-15)	LD	H	URBAN	Teacher	8.83	13% (Quality)			
	MBD	6	SW	12 (17-18)	LD	AA (n = 1) C (n = 5)	SUB	RES	8.83	44% (Quality) 64% (Production)	70% (Quality) 17% (Production)	50% (Quality) 0% (Production)	38% (Quality) Setting 58% (Production)
Zipprich (1995)	MBD	3	SW	3 to 5 (9-10)	LD	C (n = 2) H (n = 1)	NR	Teacher	7.83	59% (Quality) 9% (Production)			
Martin & Manno (1995)	MBD	3	SW	7 (13)	LD	AA (n = 1) C (n = 1) AA & C (n = 1)	URBAN	RES	6.50	75% (Elements)			
Beals (1983)	MBD	9	FR	10 (15-16)	LD (n = 3)	AA (n = 4) C (n = 5)	URBAN	Teacher	7.75	100% (Complete Sentences) 89% (Complicated Sentences)	88% (Complete Sentences) 80% (Complicated Sentences)		
	MBD	3	SW	— (11-13)	ED	C (n = 1) H (n = 1) H&AA (n = 1)	NR	Teacher	4.00	80% (Complete Sentences)			
Johnson (2005)	MBD	36	FR	7 to 8 (11-13)	Autistic MMR LD VI OHI	C (n = 11) H (n = 25)	NR	Teacher	9.00	83% (Complete Sentences)	89% (Complete Sentences)		
Schmidt et al. (1988)	MBD	7	SW	10 to 12 (14-16)	LD	C	SUB	Teacher	7.00	78% (Complete Sentences)	50% (Complete Sentences)		
Eads (1991)	MBD	9	SW	6 to 7 (12-14)	LD	NR	RURAL	Teacher	5.67	89% (Complete Sentences)			
Moran et al. (1981, study 1)	MBD	3	SW	8 to 9 (14-16)	LD	NR	SUB	Teacher	6.67	100% (Elements)	100% (Elements)		100% (Elements) Setting
Moran et al. (1981, study 2)	MBD	5	SW	8 to 9 (13-16)	LD	NR	SUB	Teacher	7.70	100% (Elements)			
Sonntag & McLaughlin (1984)	MBD	6	FR	8 to 9	—	NR	RURAL	Teacher	5.40	100% (Elements)			
Wallace & Bott (1989)	MBD	4	SW	8	LD	NR	RURAL	RES	6.67	89% (Elements)			100% (Elements) Setting

Table 2 (continued)

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Instr.	Quality score (0–11)	PND TX	PND POST	PND MAINT	PND GEN
Dowell et al. (1994)	MBD	3	SW	9 to 11 (14–16)	LD	NR	URBAN	Teacher	5.58	34% (Production) 4% (Quality)			
Goddard (1998)	ABAB	7	SW	5 to 6 (10–12)	LD	C	Self-monitoring SUB RES		9.83	29% (Production)			15% (Vocab.)
Harris et al. (1994)	MBD	4	SW	5 to 6 (10–12)	LD	AA (n = 1) C (n = 3)	SUB	Teacher	7.83	43% (Production)			
Jackson (1994)	MBD	6	SW	- (11–13)	—	NR	NR	Teacher	6.40	82% ^a (Production) 42% (Vocab)			
Moran (2004)	MBD	4	SW	2 (7–8)	ADHD (n = 2)	AA (n = 1) AI (n = 1) C (n = 2)	SUB	RES	11.00	23% (Production)			
Rumsey & Ballard (1985)	ABAB	7	SW	— (9–11)	—	NR	NR	Teacher	5.90	52% ^a (Production)			
Shimabukuro (1999)	MBD	3	SW	6 to 7 (12–13)	ADHD & LD	NR	NR	Teacher	3.67	91% ^a (Production) 75% (Accuracy)			
Wolfe (1997)	ABAB	4	SW	2 to 3 (9)	LD	NR	URBAN	RES	9.50	37% (Production)			
Montague & Leavell (1994)	MBD	9	SW	7 to 8 (12–14)	LD	AA (n = 1) C (n = 5) H (n = 3)	URBAN	RES	5.33	29% (Quality)			

Note. Dashes indicate that data were not reported. ATD = alternating treatment design; ABAB = withdrawal design; CCD = changing criteria design; MBD = multiple-baseline design; N = Number of participants on single subject graph(s); AVG = average; ESOL = English as a second language; NR = not reported; SW = struggling writer; FR = full range; ADHD = attention-deficit/hyperactivity disorder; BD = behavioral disorder; EBD = emotional and behavioral disorder; ED = emotionally disturbed; LD = learning disability; MLT = mild language delay; MMR = mild mental retardation; OHI = other health impairment; OI = orthopedic impairment; SLI = speech or language impairment; VI = visual impairment; 504 = student plan with goals and objectives related to writing; AA = African American; AI = American Indian; C = Caucasian; H = Hispanic; SUB = suburban. Instr. = instructor; RES = researcher; PND = percentage of nonoverlapping data; TX = during treatment; POST = immediately after treatment (less than 3 weeks after treatment ended); MAINT = maintenance (3 or more weeks after treatment ended); GEN = generalization (generalization to other settings or genres); Elements = basic parts of a genre or type of writing; Production = included number of words, ideas, and/or T-units, as well as number of assignments completed; Vocab = included number of different types of words.

^a Experimental control established.

Table 3
Summary PND Statistics for Writing Treatments With Four or More Studies

Treatment category	Measure	N	Grade range	Writer type	M	Mdn	Range
Strategy instruction: Planning/drafting	Elements	21	2–8	FR, SW	96% TX/POST	100% TX/POST	67%–100% TX/POST
	Elements	18	2–8	FR, SW	90% MAINT	100% MAINT	25%–100% MAINT
	Elements	4	2–3, 5–6	SW	85% GEN Genre	86% GEN Genre	67%–100% GEN Genre
	Production	10	2–8	FR, SW	91% TX/POST	95% TX/POST	68%–100% TX/POST
	Production	7	2–5, 7–8	FR, SW	86% MAINT	100% MAINT	33%–100% MAINT
Teaching grammar/usage	Quality	5	2–4, 7–8	FR, SW	99% POST	100% POST	97%–100% POST
	Grammar	4	2, 5–6	SW	83% TX	84% TX	75%–88% TX
Goal setting for productivity	Production	7	2–5, 8–12	FR, SW	79% TX/POST	91% TX	26%–100% TX
Strategy instruction: Editing	Errors	5	4, 8–12	FR, SW	84% POST	100% POST	50%–100% POST
	Corrected						
Word processing	Production	4	1, 4–5	SW	70% TX	75% TX	29%–100% TX
Reinforcement	Production	4	3–6	FR, SW	96% TX	100% TX	84%–100% TX
Prewriting activities	Quality	4	3–5, 8, 12	SW	52% TX/POST	55% TX/POST	13%–84% TX/POST
Sentence construction	Complete sentences	5	6–8, 10–12	FR, SW	86% TX/POST	83% TX/POST	78%–100% TX/POST
Strategy instruction: Paragraph construction	Elements	4	8–9	FR, SW	97% TX/POST	100% TX/POST	89%–100% TX/POST
	Self-monitoring	Production	7	2, 3, 5–7	SW	51% TX	43% TX

Note. Production includes number of words, ideas, and/or T-units, as well as number of assignments completed. N = number of studies; FR = full range; SW = struggling writer; PND = percentage of nonoverlapping data points; TX = scores during treatment; TX/POST = scores during treatment as well as immediately after treatment (less than 3 weeks after treatment ended); MAINT = maintenance (3 or more weeks after treatment ended); GEN = scores for generalization probes.

physical location of the study was only offered about one third of the time (these findings raise concerns about external validity). Likewise, fidelity of treatment was established in less than one half of the studies (possibly because this has not been stressed until recently) but adequate description of testing procedures and the establishment of social validity of the treatment occurred in about two third of the studies.

The findings for each of the 10 treatments in Table 1 are presented next. Each treatment is presented separately using a common format. First, information on studies testing the treatment is summarized. This is followed by the summary PND statistics (mean, median, and range) and interpreting these statistics in light of the quality of the studies on which they were based (the 11 quality indicators in Table 5 and the establishment of experimental control; see Table 2).

Strategy Instruction: Planning/Drafting

Twenty-five studies examined the effectiveness of teaching strategies for planning/drafting specific types of text (see Table 2). As noted earlier, all of these investigations used the SRSD model (Harris & Graham, 1996) to teach these strategies. A common component in most of the planning/drafting strategies taught was that students used specific features of the target genre (e.g., characters, location, time, goals of characters, actions, ending, and reactions for stories) to help them generate and organize possible writing ideas. Planning/drafting strategies focused on stories (e.g., Adkins, 2005; Albertson, 1998; Albertson & Billingsley, 1997), persuasive text (e.g., De La Paz & Graham, 1997; Lienemann, 2006; Mason & Shriner, 2008), and expository essays (e.g., De La

Table 4
Summary PND Statistics for Strategy Instruction Using Planning/Drafting for Stories and Expository Writing

Measure	N	Grade range	Writer type	M	Mdn	Range
Stories						
Elements	13	2–8	FR, SW	97% TX/POST	100% TX/POST	67%–100% TX/POST
Elements	9	2–5	FR, SW	93% MAINT	100% MAINT	67%–100% MAINT
Production	7	2–8	FR, SW	91% TX/POST	96% TX/POST	68%–100% TX/POST
Production	4	2–5	SW	79% MAINT	92% MAINT	33%–100% MAINT
Expository						
Elements	8	2–8	FR, SW	94% POST	100% POST	68%–100% POST
Elements	8	2–8	FR, SW	89% MAINT	100% MAINT	25%–100% MAINT

Note. Elements = basic parts of a genre or type of writing; Production includes number of words, ideas, and/or T-units, as well as number of assignments completed. N: number of studies; FR = full range; SW = struggling writer; PND = percentage of nonoverlapping data points; TX = scores during treatment; TX/POST = scores during treatment as well as immediately after treatment (less than 3 weeks after treatment ended); MAINT = maintenance (3 or more weeks after treatment ended).

Table 5
Quality Means, Standard Deviations, and Percentages for Writing Intervention Studies Based on Treatment Type

Treatment	Total quality score		N	Participant description	Participant selection	Physical description	DV(s) Quantified	DV(s) Reliable	Multiple baseline data points	Multiple interven. data points	Treatment description	Fidelity of treatment	Testing procedure description	Social validity
	M	SD												
All studies	7.90	1.70	76	51.32	47.37	30.26	94.74	90.79	92.11	90.79	93.42	43.42	68.42	59.21
Strategy instruction plan/draft	8.87	1.19	25	68.00	52.00	12.00	96.00	92.00	96.00	96.00	100.00	84.00	84.00	80.00
Teaching grammar/usage	8.18	0.77	4	50.00	50.00	50.00	100.00	100.00	100.00	100.00	100.00	0.00	75.00	75.00
Goal setting for productivity	8.16	1.03	7	28.57	28.57	42.86	100.00	100.00	100.00	100.00	85.71	42.86	85.71	42.86
Strategy instruction editing	7.94	1.76	5	40.00	80.00	40.00	80.00	100.00	60.00	80.00	100.00	40.00	20.00	80.00
Word processing	7.36	2.10	5	60.00	80.00	40.00	100.00	80.00	80.00	100.00	100.00	0.00	60.00	40.00
Reinforcement	6.99	1.23	7	14.29	14.29	28.57	100.00	100.00	100.00	100.00	100.00	0.00	85.71	28.57
Prewriting activities	8.00	1.10	5	100.00	40.00	60.00	100.00	100.00	100.00	80.00	100.00	0.00	100.00	40.00
Sentence construction	6.68	1.93	5	60.00	40.00	20.00	100.00	60.00	100.00	60.00	80.00	40.00	20.00	60.00
Strategy instruction paragraph construction	6.40	0.94	5	0.00	40.00	40.00	100.00	100.00	60.00	80.00	60.00	20.00	40.00	40.00
Self-monitoring	7.43	2.53	8	50.00	50.00	37.50	75.00	62.50	100.00	100.00	87.50	50.00	50.00	50.00

Note. Maximum number of quality points = 11. N = number of studies; DV = dependent variable; Interv. = intervention.

Paz, 1999, 2001; Mason, Snyder, Sukhram, & Kedem, 2006), and they are grouped in this fashion in Tables 2 and 3.

Twenty-one of the 25 studies had a common outcome measure: elements. These structural elements included the basic parts of a composition that students generated ideas for when planning. In 18 of the 21 studies that assessed elements, maintenance effects (3 or more weeks after the termination of treatment) for elements were assessed. In addition four studies examined generalization of elements from one genre to another (i.e., story to personal narrative). Separate PNDs (mean, median, and range) were computed for each of these conditions.

There were 108 students (almost equally divided between boys and girls) in the 21 planning/drafting intervention studies that included a measure of elements (see Table 2). These youngsters were in Grades 2–8, and most of them were struggling writers (18 of the 21 studies). There were a variety of students with disabilities, but the most common group was children with learning disabilities (other students with disabilities included students with ADHD, emotional and behavioral disorders, mild language delay or speech/language impairment, and mild mental retardation). In one study (Albertson & Bilingsley, 1997), gifted students were the target of instruction (typical writers were the focus in the other two studies). Race/ethnicity was reported in 15 studies, and the majority of students were Caucasian (55%; 48/87), followed by African American (37%; 32/87), Arab (5%; 4/87), Hispanic (2%; 2/87), and Asian (1%; 1/87). When locale was reported ($n = 16$ studies), suburban was the most common location (56%; 9/16), followed by rural (25%; 4/16) and urban settings (19%; 3/16). Researcher most commonly delivered instruction (81%; 17/21), and all of the studies employed multiple-baseline designs. It is also important to note that the 18 studies that assessed maintenance of elements had the same basic characteristics described above, but the four studies that assessed generalization of elements to a different genre only included struggling writers in Grades 2, 3, 5, and 6.

Teaching students a planning/drafting strategy had a large impact on increasing the number of basic genre elements in their writing, and this effect maintained over time (see Table 3). Median PND for treatment and maintenance was 100% in both cases. Mean PND was 96% (range = 100% to 67%) and 90% (range = 100% to 25%), respectively. Teaching these strategies also had a moderate impact on enhancing the generalization of elements from an instructed genre to an uninstructed one. Median PND was 86%, and mean PND was 85% (range = 100% to 67%).

Ten of the 25 planning/drafting strategy instructional studies assessed productivity (i.e., number of words written) during or immediately after instruction (mostly struggling writers in Grades 2–8 but also typical writers in Grades 6–8), with 7 of these studies assessing maintenance of productivity (focusing almost exclusively on struggling writers in Grades 2–8, with some typical writers in Grades 7 and 8). Moreover, 5 of the 25 studies assessed and graphed quality of students writing immediately following instruction (this involved students in Grades 2–4 and Grades 7 and 8 with a mixture of struggling and typical writers).

Teaching students a planning/drafting strategy had a large impact on productivity and quality during and/or immediately following instruction (see Table 3). Median PND for productivity and quality were 95% and 99%, respectively. Mean PND was 91% (range = 100% to 68%) and 99% (range = 100% to 97%), respectively. In addition, students generally maintained productiv-

ity gains, as median and mean PND were 100% and 86%, respectively (range = 100% to 33%).

We also examined the impact of strategy instruction for planning/drafting on writing genres. We computed separate PNDs for stories and expository writing (we combined persuasive text and expository essays for this second category). As can be seen in Table 4, this treatment had a large effect on both genres in terms of elements and productivity (all median and mean PNDs were 91% or greater). The only exceptions involved maintenance of elements for expository writing (mean PND = 89%) and maintenance of productivity for stories (mean PND = 79%).

The studies examining the teaching of strategies for planning/drafting were of high quality (on average, 82% of the quality indicators were met in all studies; see Table 5). The only quality indicator that was not met in at least 50% of the studies was description of the setting where instruction took place. Most importantly, experimental control was established in all but three of the studies that were used to compute summary PNDs for elements, productivity, or quality.

In summary, strategy instruction was effective in enhancing the number of elements, written output, and quality of students' writing, and the effects for elements and productivity were maintained over time. In addition, when transfer to a different genre was measured, positive effects were also obtained. For the most part, these findings generalize best to struggling writers in Grades 2–8 but are also valid for typical writers in Grades 4–8. Considerable confidence can be placed in these findings, as studies were of high quality and experimental control was established in almost all of them. The only concern involves failure for all but a few of the studies to adequately describe the setting where instruction took place. This narrows any generalization about conditions under which this treatment is effective.

Teaching Grammar/Usage

Four studies evaluated the effectiveness of teaching grammar/usage (see Table 2). Teaching grammar/usage included peers directly teaching capitalization skills to classmates (Campbell, Brady, & Linehan, 1991) to teachers directly teaching adverbial phrases and possessives (Dowis, 1991; Dowis & Schloss, 1992) to how to correct capitalization, subject/verb agreements, conjunctions, incomplete sentences, and run-on sentences (Hermann, Semb, & Hopkins, 1976). All four studies included a similar outcome measure that focused on correct use of grammar. In all instances, the grammar measure assessed grammar skills students were taught during treatment.

There were 30 students in the four grammar/usage studies (see Table 2). The participants were all struggling writers in Grades 2, 5, and 6. In all but one of the studies (Hermann et al., 1976), the students were either learning disabled or classified as having a mild mental retardation (1 child). In the three studies that provided information on gender, boys and girls were represented almost equally.

In the two studies that provided information on race/ethnicity, most students were Caucasians (71%; 5/7) followed by African Americans (29%; 2/7). With the exception of the one study where peers delivered instruction (Campbell et al., 1991), teachers were the instructors. The effectiveness of grammar/usage instruction was tested via multiple-baseline and withdrawal designs.

Directly teaching grammar/usage had a moderate effect on improving grammar skills (see Table 3). The median PND was 84%, whereas the mean PND was 83% (range = 88% to 75%). On average, the four studies in this category met approximately 75% of the 11 quality indicators (see Table 5). None of the studies established treatment fidelity, but experimental control was established in three of the four investigations (see Table 2).

In summary, directly teaching grammar/usage had a positive impact on the grammar skills of struggling writers in Grades 2, 5, and 6. Studies were generally of high quality, but there was uncertainty that the treatment was delivered as intended in any of the studies.

Goal Setting for Productivity

Seven studies were located that examined the impact of setting goals (see Table 2). Goal setting in these studies ranged from teachers encouraging students to exceed their previous publicly posted writing performance and receiving immediate feedback on their success (Van Houten, Morrison, Jarvis, & McDonald, 1974) to teachers setting a goal for how much students would write, with students monitoring their success in meeting the goal, and placing a star on a public chart if the goal was met (Wolfe, 1997). All of the studies included a common outcome measure: productivity. Productivity focused on amount of text produced and included the following measures: (a) number of words (Hopman & Glynn, 1989; Van Houten & McKillop, 1977; Van Houten et al., 1974; Wolfe, 1997), (b) number of words plus number of adjectives (Weygant, 1981), (c) number of writing lessons completed (Seabough & Schumaker, 1994), and (d) percentage of writing assignments completed (Kastelen, Nickel, & McLaughlin, 1984).

The seven goal setting studies included 129 students in Grades 2–5 and 8–12 (see Table 2). In the four studies that provided information on gender, the majority of participants were boys (85%, 17/20). Only one study reported any information on race/ethnicity (i.e., Weygant, 1981; all participants were Caucasian). The participants in four studies were youngsters with learning disabilities or other struggling writers, whereas youngsters in the remaining investigations represented a full range of abilities. Two of the studies were conducted in urban schools; the other five studies did not provide information on locale. In four studies, the instructor was the students' teacher, while researchers provided instruction in the other three studies. A variety of different designs were used to assess the effectiveness of goal setting for productivity, including changing criterion, multiple-baseline, and withdrawal designs.

Depending on which measure of central tendency, mean or median, was used to summarize the overall impact of the treatment, goal setting for productivity had a large to moderate effect (see Table 3). The median PND was 91%, whereas the mean PND was 79% (range = 100% to 26%).

On average, the seven studies in this category met approximately 75% of the 11 quality indicators (see Table 5). Less than one half of the studies, however, adequately described participants, selection of participants, or the physical setting for the treatment. A similar problem occurred in terms of establishing fidelity of treatment or collecting data on social validity. In addition, only three of the seven studies achieved experimental control (see Table 2).

In summary, goal setting had a large to moderate effect on increasing writing productivity of students in Grades 2–5 and 8–12 (this includes struggling and regular writers). Although studies were generally of high quality, just three of the seven studies established experimental control. Furthermore, failure to adequately describe participants, settings, and selection procedures in a majority of studies narrows generalizations about whom and under what conditions this treatment is effective. There was also uncertainty, at least in some studies, that the treatment was delivered as intended, and it was unclear whether teachers and students valued the treatment.

Strategy Instruction: Editing

Five studies examined the effectiveness of strategy instruction for editing (see Table 2). The editing strategies used in these studies ranged from an editing strategy developed as part of the University of Kansas strategy curriculum (Beals, 1983; Schmidt, 1983; Schumaker et al., 1982; Scott, 1993) to an editing strategy designed to help students use spell checkers when editing (McNaughton, Hughes, & Ofiesh, 1997). All of the studies included a common outcome measure: errors corrected. Errors corrected ranged from spelling errors corrected (McNaughton et al., 1997) to the correction of a broad range of errors (e.g., spelling, usage, and punctuation).

There were 32 participants in the five editing strategy instruction intervention studies (see Table 2). These youngsters were in Grade 4 and Grades 8–12. Each of the studies included struggling writers with learning disabilities, but one study primarily involved students representing the full range of writing abilities (Beals, 1983). Most of the participants were boys (72%; 23/32). When race/ethnicity was reported (three studies), the majority of students were Caucasian (65%; 13/20), followed by African Americans (35%; 7/20). Studies occurred in a variety of locations (two urban, one suburban, one rural, and one not reported). Teachers taught the editing strategy in 60% of the studies, and effectiveness was evaluated in all studies using multiple-baseline designs.

Depending on which measure of central tendency, median or mean, was used to summarize the overall impact of the treatment, teaching an editing strategy had a large to moderate effect on correcting errors in writing (see Table 3). The median PND was 100%, whereas the mean PND was 84% (range = 100% to 50%).

The researchers met most of the quality indicators in studies examining editing strategy instruction (72%), but more than one half of the studies failed to adequately describe participants or the setting where instruction took place (see Table 5). Testing procedures were not adequately described in a majority of studies, nor was treatment fidelity. Experimental control was only achieved in two of the five studies (see Table 2).

In summary, teaching an editing strategy had a large to moderate impact on decreasing errors in the writing of students in Grade 4 and Grades 8–12 (struggling writers and regular writers). Although studies were generally of high quality, some caution must be applied in interpreting this finding, as only two studies (out of five) established experimental control. In addition, failure to adequately describe participants and settings in a majority of studies narrows generalizations about whom and under what conditions this treatment is effective. There was also some uncertainty in a

majority of the studies that the treatment was delivered as intended.

Word Processing

Five studies evaluated the effectiveness of word processing (see Table 2). Word processing in these studies ranged from students using word processing to write versus handwriting (Brigman, 1994; Burnett, 1984; Christensen, 1993; Handley-More et al., 2003) to students using word processing with a reminder to use the editing features of word processing software (Jones, 1985). Four of the studies used a similar outcome measure: productivity. Productivity included (a) total number of words (Christensen, 1993; Jones, 1985), (b) total number of words used per session (Brigman, 1994), and (c) total number of words per minute (Handley-More et al., 2003).

There were 27 students in the four word processing studies that assessed writing productivity. All of the students were struggling writers in Grades 1, 4, and 5. In one study, these struggling writers were learning disabled (Handley-More et al., 2003), and in another investigation, they were gifted underachievers (Christensen, 1993). Two of every 3 students in these four studies were male, and the studies occurred in urban, suburban, and rural locations. Race/ethnicity was reported in two studies, and the majority of these participants were Caucasian (83%; 15/18) followed by African Americans (17%; 3/18). Teachers were the instructors in two studies, and researchers were the instructors in the other two investigations. The effectiveness of word processing on productivity was tested via alternating treatment, multiple-baseline, and withdrawal designs.

Word processing had a moderate effect on increasing students' productivity (see Table 3). The median PND was 75%, whereas the mean PND was 70% (range = 100% to 29%). On average, the four studies in this category met close to 70% of the 11 quality indicators (see Table 5). None of the studies, however, established treatment fidelity, and less than one half of them adequately described the setting where instruction took place or assessed social validity. Experimental control was established in one half of the studies.

In summary, word processing had a positive impact on increasing productivity of struggling writers in Grades 1, 4, and 5. Some caution must be exercised in interpreting this finding, as experimental control was only established in 50% of studies. Failure to adequately describe the setting in more than one half of studies narrows any generalization about the conditions under which this treatment is effective. There was also uncertainty in all of the studies whether the treatment was delivered as intended. It was unclear whether teachers or students valued this treatment.

Reinforcement

Seven studies examined the effectiveness of using reinforcement to enhance writing performance (see Table 2). Reinforcement in these studies included teacher praise for using different parts of speech in stories (Blase-Maloney, Jacobson, & Hopkins, 1975), obtaining reinforcement through public posting of writing performance (Ballard & Glynn, 1975), and receiving group reinforcement for including a certain number of words, different words, and new words in compositions (Brigham, Graubard, & Stans, 1972).

Four of the studies included a common outcome measure: productivity. Productivity measures in these studies included number of sentences, number of words, number of different types of words, and number of different sentence beginnings produced during writing (Ballard & Glynn, 1975; Blase-Maloney & Hopkins, 1973; Blase-Maloney et al., 1975; Brigham et al., 1972).

There were 60 students in Grades 3–6 in the four studies examining reinforcement (see Table 2). The participants' disability status, gender, and race/ethnicity were not reported in any study. Only two of the studies provided information on type of writer (Brigham et al., 1972, worked with struggling writers, whereas students in Ballard and Glynn, 1975, represented the full range of writing abilities). Likewise, only two studies specified location (Ballard & Glynn, 1975, were in a rural location, while Blase-Maloney et al., 1975, were in an urban school). Teachers delivered the treatment, and multiple-baseline designs were used in all studies.

Reinforcement had a large effect on students' writing productivity (see Table 3); the median and mean PND were 100% and 96%, respectively (range = 100% to 84%). The overall quality of reinforcement studies was not strong (see Table 5). None of the studies established treatment fidelity, and less than one in three studies collected social validity information. In addition, adequate descriptions of participants, settings, or selection processes were rare. Finally, only two of the four reinforcement studies that assessed productivity established experimental control (see Table 2).

In summary, reinforcement had a large impact on increasing the writing productivity of students in Grades 3–6 (both struggling and regular writers). Caution must be applied in interpreting this finding, as studies were generally of low quality, and only two (out of four) studies that assessed productivity established experimental control. Failure to adequately describe participants, settings, and selection procedures in a majority of studies narrows generalizations about whom and under what conditions this treatment is effective. There was uncertainty that the treatment was delivered as intended, and it was unclear whether teachers and students valued it.

Prewriting Activities

The effectiveness of five prewriting activities was examined in four publications. The prewriting activities included using a computer prewriting outline to generate and organize information (Channon, 2004), learning to use a graphic organizer for generating ideas prior to persuasive writing (Thanhouser, 1994), and learning to use a story web for generating ideas prior to writing (Zipprich, 1995). Four prewriting comparisons in three studies (Channon, 2004; Thanhouser, 1994; Zipprich, 1995) included a common measure: writing quality. Quality was assessed using a holistic measure (see Graham & Perin, 2007b).

There were 16 students in the studies that examined the impact of prewriting strategies on the quality of writing (see Table 2). Participants were in Grades 3–5 (Zipprich, 1995), 8 (Channon, 2004), and 12 (Thanhouser, 1994). All of the students in these studies were struggling writers with learning disabilities. The majority of the youngsters in these studies were male (94%; 15/16) and Hispanic (50%; 8/16). The remaining students were Caucasian (44%; 7/16) or African Americans (6%; 1/16). In all but one case, the intervention was delivered by the students' teacher. Two of the

three studies indicated geographic location, which were urban and suburban schools, respectively. The effectiveness of prewriting in the three studies was assessed via alternating treatment and multiple-baseline designs.

The median and mean PND for prewriting were similar: 55% and 52%, respectively (range = 84% to 13%). Thus, prewriting had a small effect on improving writing quality.

The researchers met most of the quality indicators in studies examining prewriting activities (73%), but less than one half of the studies adequately described participant selection, established treatment fidelity (this was not done in any study), or assessed social validity (see Table 5). Experimental control was only achieved in one of the four comparisons that assessed writing quality (see Table 2).

In summary, prewriting activities had a small impact on improving the quality of writing produced by struggling writers in Grades 3–5, 8, and 12. Although studies were generally of high quality, caution must be applied in interpreting this finding, as only one comparison (out of four) established experimental control. In addition, the value of the treatment to teachers and students was unclear, and there was uncertainty that the treatment was delivered as intended.

Sentence Construction

Five studies examined the effectiveness of teaching sentence construction skills (see Table 2). Sentence construction in these studies included teaching strategies for writing different types of sentences (Beals, 1983; Eads, 1991; Johnson, 2005; Schmidt, Deshler, Schumaker, & Alley, 1988; these strategies were developed as part of the University of Kansas strategy curriculum) and teaching students how to combine simpler sentences into more complex ones (First, 1994). All of the studies included a common measure: complete sentences. Complete sentences were measured by calculating the percentage of complete sentences independently written by students.

There were 64 participants in the five sentence construction studies (see Table 2). These youngsters were in Grades 6–8 and 10–12, and the studies included more boys than girls (66%; 42/64). In all studies, there were students with disabilities (mostly learning), but two studies focused mainly on students representing the full range of writing ability (Beals, 1983; Johnson, 2005). Four studies provided information on race/ethnicity; the majority of students were Hispanic (47%; 26/55), followed by Caucasian (44%; 24/55), African American (7%; 4/55), and Hispanic and African American (2%, 1/55). In the three studies that specified locale, they were evenly split between urban, suburban, and rural locations. In all cases, instruction was delivered by teachers, and multiple-baseline designs were used to test the effectiveness of teaching sentence skills.

Teaching sentence construction was an effective treatment (see Table 3). The median PND was 83%, whereas the mean PND was 86% (range = 100% to 78%). However, none of the studies established experimental control (see Table 2), and the overall quality of studies was poor (on average, each study only met 61% of the 11 quality indicators; see Table 5). More than one half of the studies did not meet the following criteria: participant selection, physical description of setting, fidelity of treatment, and testing procedure description.

In summary, sentence construction was an effective practice in increasing the percentage of complete sentences produced by students in Grades 6–8 and 10–12 (for struggling and regular writers). Caution must be exercised in interpreting this finding, as the research in this area was particularly weak.

Strategy Instruction: Paragraph Construction

Five studies examined the effectiveness of teaching students strategies for constructing paragraphs. These five studies were included in four publications (Dowell, Storey, & Gleason, 1994; Moran, Schumaker, & Vetter, 1981; Sonntag & McLaughlin, 1984; Wallace & Bott, 1989). Paragraph construction strategies ranged from teaching students strategies for organizing and/or writing expository paragraphs (Moran et al., 1981; Sonntag & McLaughlin, 1984; Wallace & Bott, 1989) to teaching students how to write descriptive paragraphs (Dowell et al., 1994). A conceptually similar outcome measure, writing elements, was used in four of the examinations. Writing elements for paragraphs involved determining whether basic parts of a paragraph (e.g., topic sentence and concluding sentence) were evident and/or correctly used.

There were 18 students in Grades 8–9 in the four strategy instruction for paragraph studies where elements were assessed (see Table 2). Two thirds of these students were struggling writers with learning disabilities, whereas the remaining students represented the full range of writing ability. In the three studies where gender was reported, 67% of students were boys. No information on race/ethnicity was provided in any study. Two of the studies took place in a suburban location, with one study conducted in an urban school (no location was given for the fourth study). In all instances but one study (Wallace & Bott, 1989), teachers taught the strategies. Only multiple-baseline designs were used to test the effectiveness of this treatment.

Teaching strategies for writing paragraphs had a large and positive impact on the schematic structure (i.e., elements) of students' paragraphs (see Table 3). The median PND was 100%, whereas the mean PND was 97% (range = 100% to 89%). However, studies were of poor quality (on average, only 58% of the 11 quality indicators were met; see Table 5). More than one half of the studies did not meet the following criteria: participant description, participant selection, physical description of treatment setting, treatment fidelity, testing procedure description, or social validity. None of the studies established experimental control.

In summary, teaching a paragraph strategy had a large effect on the paragraph writing of students in Grades 8 and 9 (struggling and regular writers). Considerable caution must be exercised in interpreting this finding, as none of the studies established experimental control. Moreover, failure to adequately describe participants, selection procedures, and settings in a majority of studies narrows generalizations about whom and under what conditions this treatment is effective. It was unclear whether the treatment was delivered as intended in the majority of studies, nor was it clear that teachers or students valued this treatment.

Self-Monitoring

Eight studies examined the effects of self-monitoring procedures (see Table 2). Self-monitoring procedures in these studies

included monitoring and graphing: the number of words and sentences written (Goddard, 1998), on-task behavior (Harris, Graham, Reid, McElroy, & Hamby, 1994; Moran, 2004; Rumsey & Ballard, 1985), or on-task behavior plus written productivity (Wolfe, 1997). Seven studies included a similar measure: productivity. Productivity included (a) total number of sentences (Goddard, 1998), (b) total number of words (Goddard, 1998; Harris et al., 1994; Jackson, 1994; Moran, 2004; Rumsey & Ballard, 1985; Wolfe, 1997), and (c) total number of written items completed (Shimabukuro, Prater, Jenkins, & Edelen-Smith, 1999).

There were 35 students in the seven self-monitoring studies (Grades 2, 3, 5–7) that assessed productivity (see Table 2). All of these students were struggling writers, and the majority of youngsters (57%; 15/35) were classified as having learning disabilities or ADHD. Most of the students were boys (80%; 28/35). Race/ethnicity was only reported in three studies, and location was provided in four of them (all but one of these studies were in suburban schools). Teachers delivered the treatment in the majority of the investigations (57%; 4/7). The effectiveness of self-monitoring was assessed through multiple-baseline and withdrawal designs.

Depending on the central tendency measure used, mean or median, self-monitoring was either ineffective or had a small impact on writing productivity (see Table 3). The median PND was 43%, whereas the mean PND was 51% (range = 93% to 23%).

On average, the self-monitoring studies met two thirds of the 11 quality indicators (see Table 5). Six quality indicators were met by just one half or less of the studies. These included participant description, participant selection, physical description of setting, treatment fidelity, description of testing procedures, and social validity. In addition, less than one half of the seven self-monitoring studies that assessed productivity established experimental control (see Table 2).

In summary, the effectiveness of self-monitoring was unclear, as the mean PND was in the small effect range, but the median PND indicated that the treatment was ineffective. Moreover, considerable caution must be exercised in interpreting any conclusions for this treatment, as experimental control was established in only 43% of studies. Failure to adequately describe participants, settings, and selection procedures in the majority of studies narrows generalizations about whom and under what conditions this treatment is effective. In addition, only one half of the studies established treatment fidelity, weakening the claim that the independent variable was responsible for changes in students' writing behaviors.

Discussion

Writing is a critical skill in an advanced technological society. The future aspirations of students who do not master the writing process are at risk, as strong writing skills are needed to attend college and obtain more than menial employment (Graham & Perin, 2007b). An important ingredient in ensuring that students become skilled writers involves teachers' use of effective writing practices. Previous meta-analyses have examined true- and quasi-experimental writing intervention studies to identify such practices (Bangert-Drowns, 1993; Bangert-Drowns et al., 2004; Goldring et al., 2003; Graham, 2006a; Graham & Harris, 2003; Graham &

Olinghouse, in press; Graham & Perin, 2007a, 2007b; Hillocks, 1986). This meta-analysis extends these efforts to single subject design writing intervention studies, providing a broader base for making evidence-based recommendations for teaching writing.

Caveats and Limitations

Before summarizing the primary findings from this review, it is important to explore the limitations of such an analysis. First, a meta-analysis, like the one conducted here, involves aggregating the findings from individual studies to draw a more general conclusion about a treatment. The value and breadth of a general conclusion depends on a variety of factors, such as who participates in the studies (grade and type of learner) and the quality of the investigations. For example, it would be inappropriate to draw a broad conclusion about a treatment for all students if the treatment was only tested with primary grade students. Consequently, our generalizations about the effectiveness of specific treatments are appropriately restricted to the grades and types of students tested. They are also constrained by study quality. Of course, more confidence can be placed in a conclusion drawn from this review if it is also supported by findings from other types of research (e.g., experimental and quasi-experimental research). Consequently, we also relate the findings from this review to prior reviews (e.g., Graham & Perin, 2007a, 2007b; Hillocks, 1986). In fact, we think that decisions concerning evidence-based teaching practices are most productive when they are based on all available evidence, including qualitative, correlational, experimental, and single subject research (Graham & Harris, in press).

Second, the effect size metric, PND, used in this study can only be calculated when data for each assessment point is available in the research report. Fortunately, we had only to eliminate two studies that did not provide the needed information (many more would have been eliminated if we had used the regression effect method). Nevertheless, it was not possible to compute PND for overall writing quality in most of the studies included in this review (this variable was often measured but not graphed). This variable has been the primary focus of previous meta-analyses of true- and quasi-experimental writing intervention research (e.g., Graham & Perin, 2007a; Hillocks, 1986). It is also important to note that PND and Cohen's *d* (the effect size of choice in the meta-analyses of group experimental studies) are not directly comparable.

Third, some instructional procedures have been the focus of more single subject design research than others. For example, strategy instruction for revising may be an effective procedure (see Appendix), but there are not enough single subject studies available to draw even a tentative conclusion about its impact. Moreover, there was only one treatment, strategy instruction for planning/drafting, that involved 10 or more single subject design tests of its effectiveness. A similar problem existed in the most current meta-analysis of true- and quasi-experimental design writing intervention studies (Graham & Perin, 2007a), as only four treatments (strategy instruction, word processing, processing writing approach, and grammar instruction) yielded 10 or more effect sizes. Less confidence can be placed in evidence-based recommendation based on a small number of studies. There is clearly a need for additional replication as well as the study of new treatments. Between this review and Graham and Perin's (2007a) meta-

analysis, only 213 investigations of writing interventions were located. It is obvious that the research base in writing is unacceptably thin. Simply put, federal agencies, such as the Institute of Educational Sciences and the National Institutes of Health, need to make writing a priority, making monies available for conducting more writing intervention research and preparing new researchers.

Fourth, one concern with meta-analysis involves comparability of outcome measures on which the effect sizes are based. We addressed this problem by only computing a summary PND (mean, median, and range) for a treatment when there were at least four or more studies that graphed a conceptually similar measure (i.e., productivity). It must be noted, however, that there were many instances where these conceptually similar measures were not exactly the same. For example, productivity typically involved number of words written but also included number of sentences. This introduces some unwanted noise into the machinery of meta-analysis.

What Instructional Practices Improve Students' Writing?

We were able to calculate summary PNDs (mean, median, and range) for 10 writing treatments. For all but one treatment, the median and mean PND are all above the ineffective range (50% or below) established by Scruggs et al. (1986). The one exception was self-monitoring, where the mean PND was slightly above 50%, but the median was below this level. We do not offer a recommendation for this treatment, as its effectiveness is questionable and research quality in this area is generally poor.

In summarizing our findings, we included a recommendation, the median and mean PND for the treatment, grade range of students tested for each treatment, and the type of student to whom the recommendation best applied (typical students, struggling writers, or both). We did not analyze the data race/ethnicity, as these statistics were absent in almost one half of the studies. We also indicated when treatment findings need to be interpreted more cautiously due to poor study quality. The 10 recommendations were ordered according to two factors. Treatments that were based on the strongest research evidence were presented before treatments based on weaker evidence (this was determined by considering both the establishment of experimental control and the quality of studies). When the strength of the evidence was generally equivalent, then treatments with larger median PNDs were presented before ones with smaller PNDs. We privileged median over mean PND, as the median is less susceptible to the influence of outliers and does not necessitate the use of an equal interval scale (PND is based on percentage).

1. Teach students strategies for planning/drafting both narrative and expository text (treatment/posttreatment median PND for elements, productivity, and quality = 100%, 95%, and 99%, respectively; mean PND = 96%, 91%, and 99%, respectively). The effects of this treatment were not only maintained over time (3 weeks or longer; median PND = 100% for both elements and productivity; mean PND = 90% and 86%, respectively) but were generalized to untaught genres (median PND for elements = 86%; mean PND = 85%). Considerable confidence can be placed in the effectiveness of this treatment with struggling writers in Grades 2–8 and typical writers in Grades 4–8. These findings also provided support for the SRSD model (Harris & Graham, 1996) of strategy instruction, as all of these studies used this model to teach the target strategies.
2. Directly teach grammar skills to struggling writers (median PND for grammar skills = 84%; mean PND = 83%). Confidence can be placed in the effectiveness of this treatment with students in Grades 2, 5, and 6.
3. Set clear and specific goals to increase students' writing productivity (median PND for productivity = 91%; mean PND = 79%). Confidence can be placed in the effectiveness of this treatment with struggling writers and typical students in Grades 2–5 and 8–12.
4. Teach students strategies for editing their compositions (median PND for reducing errors = 100%; mean PND = 84%). The confidence that can be placed in the effectiveness of this treatment with struggling and typical writers in Grade 4 and Grades 8–12 must be tempered somewhat due to concerns about the quality of the research.
5. Make it possible for students to use word processing as a primary tool for writing (median PND for productivity = 75%; mean PND = 70%). The confidence that can be placed in the effectiveness of this treatment with struggling writers in Grades 1, 4, and 5 must be tempered somewhat because of concerns about the quality of the research.
6. Reinforce students for their writing productivity (median PND for productivity = 100%; mean PND = 96%). The confidence that can be placed in the effectiveness of this treatment with struggling and typical writers in Grades 3–6 must be tempered somewhat because of concerns about the quality of the research.
7. Engage students in prewriting activities for gathering and organizing ideas in advance of writing (median PND for writing quality = 55%; mean PND = 52%). The confidence that can be placed in the effectiveness of this treatment with struggling writers in Grades 3–5, 8, and 12 must be tempered because of concerns about the lack of experimental control in the majority of studies.
8. Teach students how to form complex sentences (median PND for complete sentences produced = 83%; mean PND = 86%). The confidence that can be placed in the effectiveness of this treatment with struggling and typical writers in Grades 6–12 must be tempered somewhat because of concerns about the quality of the research and lack of experimental control.
9. Teach students strategies for writing different types of paragraphs (median PND for paragraph elements = 100%; mean PND = 97%). The confidence that can be placed in the effectiveness of this treatment with struggling and typical writers in Grades 8 and 9 must be

tempered considerably because of concerns about the quality of the research and lack of experimental control.

How the Findings From This Meta-Analysis Support, Extend, and Contradict Prior Findings

Although it is not possible to compare directly the findings from this meta-analysis and previous ones examining true- and quasi-experimental writing intervention studies, the findings from this review support and extend findings from these previous analyses. Consistent with previous reviews involving both group experimental studies (Graham, 2006a; Graham & Harris, 2003; Graham & Perin, 2007a, 2007b) and single subject design investigations (Graham, 2006a; Graham & Harris, 2003), this review provided strong support for the effectiveness of teaching planning/drafting strategies to typical and struggling writers. It also provided further validation of the effectiveness of the SRSD model (Harris & Graham, 1996). Additional work needs to be done in this area, however, as such strategic instruction has mostly involved only stories and persuasive text, it has not covered all grade levels, and there is virtually no evidence on whether its effects are maintained past 6 months.

Consistent with previous reviews of experimental group studies (Bangert-Drowns, 1993; Goldring et al., 2003; Graham & Perin, 2007a, 2007b), word processing also had a positive impact on students' writing. This analysis extends these previous findings by demonstrating the effectiveness of word processing with struggling writers (this was not specifically examined in prior reviews). Likewise, we found that prewriting activities enhanced the quality of text produced by struggling writers. A previous meta-analysis of true- and quasi-experimental studies by Graham and Perin (2007a, 2007b) found a similar impact for more typical writers.

Replications and extensions between the findings from our review and previous meta-analyses involving group experimental studies were further found for goal setting and sentence construction. Graham and Perin (2007a, 2007b) indicated that establishing clear and specific writing goals improved the quality of text produced by typical and struggling writers. Our review of single subject design studies also supports the effectiveness of this treatment with typical and struggling writers but extends its impact to writing productivity. In addition, Hillocks (1986) as well as Graham and Perin (2007a, 2007b) reported that sentence combining was an effective practice for improving the quality of students' writing (only one of the studies reviewed involved struggling writers). We found that teaching sentence construction skills (including sentence combining) had a positive impact on the number of complete sentences produced by both typical and struggling writers (although most studies involved struggling writers).

Our review of single subject design writing intervention studies provided evidence on the effectiveness of several writing treatments that were not examined in prior reviews. This included reinforcing students' writing productivity, teaching strategies for editing text, teaching strategies for constructing paragraphs, and self-monitoring of writing productivity. With the exception of self-monitoring (which involved only struggling writers and produced questionable effects), these treatments were effective with typical and struggling writers. Noticeably missing from this and other reviews (e.g., Graham & Perin, 2007a, 2007b) were recommendations specifically on teaching strategies for revising content.

Unfortunately, there are very few single subject design studies (see Appendix) or experimental studies (see Graham & Perin, 2007a) that examined the effectiveness of such instruction. Clearly, additional research is needed in this area.

One area where our results were at odds with the finding of other reviews involved the teaching of grammar/usage. In their meta-analyses of true- and quasi-experimental studies, Hillocks (1986) and Graham and Perin (2007a, 2007b) reported that traditional grammar instruction did not enhance the quality of students' writing. In a narrative review of the research in this area, Andrews et al. (2006) also indicated that such instruction did not enhance students' grammar in text. In contrast, we found that grammar/usage instruction had a positive impact on the correct use of grammar in the text written by struggling writers. One explanation for this difference involves the type of writers studied. The reviews of group experimental studies focused almost exclusively on typical writers, whereas the participants in the single subject design studies were all weaker writers. It is also possible that the difference in findings was due to the methods used to teach grammar. In the single subject design studies, grammar/usage was primarily taught by the teacher modeling how to correctly apply the skill in writing, followed by students practicing applying the skill, with teacher assistance as needed. This was not typically done in the group experimental studies. In any event, additional research is needed to determine effective procedures for improving grammar in the writing of both typical and struggling writers.

Quality of Research

This review provided some important insights into the strengths and weaknesses of current single subject design writing intervention research. When we looked at all studies together, we found that researchers consistently quantified their dependent variables, collected multiple baseline and intervention data points for writing, and adequately described their treatment. However, many researchers failed to adequately describe their participants, the processes used to select them, or the setting in which instruction took place, narrowing generalizations about to whom and under what conditions treatments are effective. In addition, evidence that a treatment was implemented as intended was provided in just 4 out of every 10 studies, raising uncertainty about treatment fidelity. It must be noted, however, that failure to collect such data, especially for older studies, does not mean that the independent variable was in fact compromised, as it has only become common to report treatment fidelity in the last 10 years or so (Graham & Perin, 2007a). Two other areas that are in need of some improvement involved adequately describing testing procedures (this did not occur in 1 out of every 3 studies) and collecting information on the social validity of a treatment (this did not occur in 4 out of every 10 studies).

Perhaps most importantly, many researchers failed to establish experimental control for one or more graphed variables in their study (this occurred 40% of the time for studies included in Table 2). There was also considerable variability across treatment categories in terms of the quality of studies and establishment of experimental control. This led us to clarify how much confidence could be placed in specific recommendations. Our analysis of the quality of individual studies indicated that there is considerable room for improvement in single subject design writing interven-

tion research, especially in terms of describing who participates and where instruction took place, documenting treatment fidelity, describing testing procedures, assessing social validity, and establishing experimental control. Researchers and those who review future research in this area (e.g., as journal reviewers and dissertation committee members) need to ensure that these problems are addressed.

Issues Involved in Implementing the Recommendations

Implementing research-based treatments is a challenging task. Just because a treatment was effective in a research study does not guarantee that it will be effective in all other situations. There is rarely, if ever, a perfect match between the conditions under which the research was implemented and the conditions in which it is subsequently put to use by practitioners. Even when the match is good, the safest course of action is to monitor continually the effects of the treatment to gauge directly whether it is effective under the new conditions.

It is also important to note that we do not know what combination or how much of each of the recommended treatments in this review or other reviews (e.g., Graham & Perin, 2007a) is needed to maximize writing instruction. There is some preliminary evidence, however, that integrating some treatments can be beneficial (see Danoff, Harris, & Graham, 1993; Sadoski, Wilson, & Norton, 1997). Furthermore, the recommendations for teaching writing from this and other recent reviews (e.g., Graham & Perin, 2007a) are incomplete, as they do not address all aspects of writing (e.g., writing vocabulary, classroom-based assessment, parental participation, and use of new technologies, motivation).

A final issue in implementing evidence-based writing practices and recommendations revolves around the different organizational structures or formats for teaching writing that exist in schools. At the elementary level, regular classroom teachers, special education teachers, other specialists (e.g., reading specialists), and aides may all be involved in one or more aspects of writing instruction. At the secondary level, writing instruction may occur within the context of the language arts or English classroom, other content classrooms (such as history or biology), or with a learning specialist (such as a special education teacher). At either level, writing might be taught as a separate subject, in conjunction with content learning in some classes or subjects, or infused throughout the curriculum. The effectiveness of these various formats has neither been tested nor compared one to another (Graham & Perin, 2007a). Furthermore, it is not certain how well the evidence-based practices identified in this or other reviews would fare in these different formats. Before implementing one or more of these treatments identified in this review, careful analysis of the organizational structure or format within which it will be placed should be undertaken, with the aim of identifying factors that may facilitate or impede effectiveness.

Concluding Comments

Meta-analysis provides a useful tool for drawing “important insight from what might otherwise be a confused and disparate literature” (Bangert-Drowns et al., 2004, p. 52). The writing intervention literature in general, and single subject design studies in this area, clearly fit this description, as they tested the effectiveness

of a wide range of interventions. We capitalized on the strengths of meta-analysis to conduct a comprehensive review of writing treatments assessed via single subject design. This was a productive strategy, as we were able to draw nine evidence-based recommendations for teaching writing.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- Achieve, Inc. (2005). *Rising to the challenge: Are high school graduates prepared for college and work?* Washington, DC: Author.
- *Adkins, M. H. (2005). *Self-regulated strategy development and generalization instruction: Effects on story writing among second and third grade students with emotional and behavioral disorders*. Unpublished doctoral dissertation, University of Maryland.
- *Albertson, L. R. (1998). *A cognitive-behavioral intervention study: Assessing the effects of instruction on story writing*. Unpublished doctoral dissertation, University of Washington.
- *Albertson, L. R., & Billingsley, F. F. (1997, March). *Improving young writers' planning and reviewing skills while story-writing*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.
- Andrews, R., Torgerson, C., Bevertson, S., Freeman, A., Locke, T., Low, G., Robinson, A., & Zhu, D. (2006). The effects of grammar teaching on writing development. *British Educational Research Journal*, 32, 39–55.
- *Ballard, K. D., & Glynn, T. (1975). Behavioral self-management in story writing with elementary school children. *Journal of Applied Behavior Analysis*, 8, 387–398.
- Bangert-Drowns, R. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63, 69–93.
- 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, 29–58.
- *Bauerschmidt, M. C. (1991). *Repeated writing to improve writing fluency for students with mild handicaps*. Unpublished doctoral dissertation, University of Wisconsin—Milwaukee.
- *Beals, V. L. (1983). *The effects of large group instruction on the acquisition of specific learning strategies by learning disabled adolescents*. Unpublished doctoral dissertation, University of Kansas.
- *Blase-Maloney, K. B., & Hopkins, B. L. (1973). The modification of sentence structure and its relationship to subjective judgments of creativity in writing. *Journal of Applied Behavior Analysis*, 6, 425–433.
- *Blase-Maloney, K. B., Jacobson, D. R., & Hopkins, B. L. (1975). An analysis of the effects of lectures, requests, teach praise and free time on the creating writing behavior of third grade children. In E. Ramp & G. Semb (Eds.), *Behavior analysis* (pp. 244–260). Englewood Cliffs, NJ: Prentice-Hall.
- *Bording, C., McLaughlin, T. R., & Williams, R. L. (1984). Effects of free time on grammar skills of adolescent handicapped students. *Journal of Educational Research*, 77, 312–318.
- *Boyer, A. W. (1990). *Improving the expository paragraph writing of learning disabled elementary school students using small group strategies instruction and word processing*. Unpublished doctoral dissertation, University of Kentucky.
- *Brigham, T. A., Graubard, P. S., & Stans, A. (1972). Analysis of the effects of sequential reinforcement contingencies on aspects of composition. *Journal of Applied Behavior Analysis*, 5, 421–429.
- *Brigman, D. J. P. (1994). *The effects of writing with pencil and paper compared with writing with a computer and a word processing program on first grade students who have writing difficulty*. Unpublished doctoral dissertation, Johns Hopkins University.

- *Burnett, J. H. (1984). *Word processing as a writing tool of an elementary school student*. Unpublished doctoral dissertation, University of Maryland.
- *Campbell, B. J., Brady, M. P., & Linehan, S. (1991). Effects of peer-mediated instruction on the acquisition and generalization of written capitalization skills. *Journal of Learning Disabilities, 24*, 6–14.
- *Campbell, J. A., & Willis, J. (1978). Modifying components of creative behavior in the natural environment. *Behavior Modification, 2*, 549–564.
- *Channon, S. (2004). *Effects of writing support programs on the writing quality and attitudes of Hispanic junior high school students with learning disabilities*. Unpublished doctoral dissertation, Illinois State University.
- *Christensen, E. E. (1993). *The effects of word processing on the creative writing of high achieving and low achieving gifted elementary students*. Unpublished doctoral dissertation, University of Washington.
- *Cole, K. B. (1992). *Efficacy and generalization of instruction in sequential expository writing for students with learning disabilities*. Unpublished doctoral dissertation, Northern Illinois University.
- *Danoff, B., Harris, K. R., & Graham, S. (1993). Incorporating strategy instruction within the writing process in the regular classroom: Effects on the writing of students with and without learning disabilities. *Journal of Reading Behavior, 25*, 295–322.
- *De La Paz, S. (1999). Self-regulated strategy instruction in regular education settings: Improving outcomes for students with and without learning disabilities. *Learning Disabilities Research & Practice, 14*, 92–106.
- *De La Paz, S. (2001). Teaching writing to students with attention deficit disorders and specific language impairment. *Journal of Educational Research, 95*, 37–47.
- *De La Paz, S., & Graham, S. (1997). Strategy instruction in planning: Effects on the writing performance and behavior of students with learning difficulties. *Exceptional Children, 63*, 167–181.
- *Dowell, H. A., Storey, K., & Gleason, M. (1994). A comparison of programs designed to improve the descriptive writing of students labeled learning disabled. *Developmental Disabilities Bulletin, 22*, 73–91.
- *Dowis, C. L. (1991). *The effects of mini-lesson instruction on the writings of students with learning disabilities within the writing process using whole group instruction*. Unpublished doctoral dissertation, University of Missouri—Columbia.
- *Dowis, C. L., & Schloss, P. (1992). The impact of mini-lessons on writing skills. *Remedial and Special Education, 13*, 34–42.
- *Eads, J. R. (1991). *Classroom teacher mediated generalization of a sentence writing strategy to the regular classroom by nine middle school students with learning disabilities*. Unpublished educational specialist thesis, Northeast Missouri State University.
- *First, C. G. (1994). *The effects of sentence combining on the written expression skills of students with serious emotional disturbances*. Unpublished doctoral dissertation, University of the Pacific.
- *Germain, J. C. (2004). *Remediation of written expression deficits in an elementary school population*. Unpublished doctoral dissertation, University of Northern Colorado.
- *Glomb, N., & West, R. P. (1990). Teaching behaviorally disordered adolescents to use self-management skills for improving the completeness, accuracy, and neatness of creative writing homework assignments. *Behavioral Disorders, 15*, 233–242.
- *Goddard, Y. L. (1998). *Effects of self-monitoring and self-evaluation on the written language performance and on-task behavior of elementary students with learning disabilities*. Unpublished doctoral dissertation, Ohio State University.
- Goldring, A., Russell, M., & Cook, A. (2003). The effects of computers on student writing: A meta-analysis of studies from 1992–2002. *Journal of Technology, Learning, and Assessment, 2*, 1–51.
- Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities. *Learning Disabilities Quarterly, 22*, 78–98.
- Graham, S. (2006a). Strategy instruction and the teaching of writing: A meta-analysis. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York: Guilford Press.
- Graham, S. (2006b). Writing. In P. Alexander & P. Wiine (Eds.), *Handbook of educational psychology* (pp. 457–477). Mahwah, NJ: Erlbaum.
- Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRS studies. In L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of research on learning disabilities* (pp. 383–402). New York: Guilford Press.
- Graham, S., & Harris, K. R. (in press). Evidence-based writing practices: Drawing recommendations from multiple sources. *British Journal of Educational Psychology* (monograph series).
- *Graham, S., & MacArthur, C. (1988). Improving learning disabled students' skills at revising essays produced on a word processor: Self-instructional strategy training. *Journal of Special Education, 22*, 133–152.
- *Graham, S., MacArthur, C., Schwartz, S., & Page-Voth, V. (1992). Improving the compositions of students with learning disabilities using a strategy involving product and process goal setting. *Exceptional Children, 58*, 322–334.
- Graham, S., & Olinghouse, N. (in press). Learning and teaching writing. In E. Anderman & L. Anderman (Eds.), *Psychology of classroom learning*. Farmington Hills, MI: Gale.
- Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445–476.
- Graham, S., & Perin, D. (2007b). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools*. New York: Carnegie Corporation of New York.
- *Guzel-Ozmen, R. (2006). The effectiveness of modified cognitive strategy instruction in writing with mildly mentally retarded Turkish students. *Exceptional Children, 72*, 281–297.
- *Handley-More, D., Deitz, J., Billingsley, F. F., & Coggins, T. E. (2003). Facilitating written work using computer word processing and word prediction. *American Journal of Occupational Therapy, 57*, 139–151.
- *Harris, K. R., & Graham, S. (1985). Improving learning disabled students' composition skills: Self-control strategy training. *Learning Disability Quarterly, 8*, 27–36.
- Harris, K. R., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation* (2nd ed.). Cambridge, MA: Brookline Books.
- *Harris, K. R., Graham, S., Reid, R., McElroy, K., & Hamby, R. S. (1994). Self-monitoring of attention versus self-monitoring of performance: Replication and cross-task comparison studies. *Learning Disability Quarterly, 17*, 121–138.
- Hayes, J. R. (2000). A new framework for understanding cognition and affect in writing. In R. Indrisano & J. R. Squire (Eds.), *Perspectives on writing: Research, theory and practice* (pp. 6–44). Newark, DE: International Reading Association.
- *Hermann, J. A., Semb, S., & Hopkins, B. L. (1976). Effects of formal "grammar" and "direct method" training on the number of errors in compositions written by sixth-graders. *Revista Mexicana de Analisis de la Conducta, 2*, 68–84.
- Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: National Council of Teachers of English.
- *Hopman, M., & Glynn, T. (1989). The effect of correspondence training on the rate and quality of written expression of four low achieving boys. *Educational Psychology, 9*, 197–213.
- Horner, R., Carr, E., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–180.
- *Jackson, H. I. (1994). *Performance feedback and generalization programming in writing instruction*. Unpublished doctoral dissertation, Syracuse University.

- *Jerram, H., Glynn, T., & Tuck, B. (1988). Responding to the message: Providing a social context for children learning to write. *Educational Psychology, 8*, 31–40.
- *Johnson, C. S. (2005). *Teaching sentence writing preskills to middle school students with mild to moderate disabilities*. Unpublished master's thesis, California State University, Fullerton.
- *Jones, M. L. D. (1985). *Effects of a microcomputer word processor on the writing and editing of fourth grade reluctant writers*. Unpublished doctoral dissertation, University of Maryland.
- *Kastelen, L., Nickel, M., & McLaughlin, T. F. (1984). A performance feedback system: Generalization of effects across tasks and time with eighth-grade English students. *Education and Treatment of Children, 7*, 141–155.
- Kennedy, C. H. (2002). The maintenance of behavior change as an indicator of social validity. *Behavior Modification, 26*, 594–604.
- *Korducki, R. A. (2001). *An instructional program integrating strategies for composition and self-regulation: Effects on the English and Spanish language writing skills of bilingual Latino students with learning difficulties*. Unpublished doctoral dissertation, University of Wisconsin—Milwaukee.
- *Kraetsch, G. A. (1981). The effects of oral instructions and training on the expansion of written language. *Learning Disability Quarterly, 4*, 82–90.
- *Lane, K. L., Harris, K. R., Graham, S., Weisenbach, J. L., Brindle, M., & Morphy, P. (2008). The effects of self-regulated strategy development on the writing performance of second-grade students with behavioral and writing difficulties. *Journal of Special Education, 41*, 234–253.
- *Li, D. (2000). *Effect of story mapping and story map questions on the story writing performance of students with learning disabilities*. Unpublished doctoral dissertation, Texas Tech University.
- *Lienemann, T. O. (2006). *Improving the writing performance of students with attention-deficit/hyperactivity disorder*. Unpublished doctoral dissertation, University of Nebraska—Lincoln.
- *Lienemann, T. O., Graham, S., Leader-Janssen, B., & Reid, R. (2006). Improving the writing performance of struggling writers in second grade. *Journal of Special Education, 40*, 66–78.
- Lipsley, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- *Martin, K. F., & Manno, C. (1995). Use of a check-off system to improve middle school students' story compositions. *Journal of Learning Disabilities, 28*, 139–149.
- *Mason, L. H., & Shriner, J. G. (2008). Self-regulated strategy development instruction for writing an opinion essay: Effects for six students with emotional/behavior disorders. *Reading & Writing: An Interdisciplinary Journal, 21*, 71–93.
- *Mason, L. H., Snyder, K. H., Sukhram, D. P., & Kedem, Y. (2006). TWA + PLANS strategies for expository reading and writing: Effects for nine fourth-grade students. *Exceptional Children, 73*, 69–89.
- *McCurdy, M. (2002). *Examining the effects of a multi-component writing program in increasing three writing skills of secondary school students*. Unpublished doctoral dissertation, Mississippi State University.
- *McGee, S. J. (1996). *The effects of goal setting and attributional feedback on self-efficacy for writing and writing achievement*. Unpublished doctoral dissertation, Florida International University.
- *McNaughton, D., Hughes, C., & Ofiesh, N. (1997). Proofreading for students with learning disabilities: Integrating computer and strategy use. *Learning Disabilities Research & Practice, 12*, 16–28.
- *Montague, M., & Leavell, A. G. (1994). Improving the narrative writing of students with learning disabilities. *Remedial and Special Education, 15*, 21–33.
- *Moran, M. R., Schumaker, J. B., & Vetter, A. F. (1981). *Teaching a paragraph organization strategy to learning disabled adolescents* (Research Rep. No. 54). Lawrence, KS: Institute for Research in Learning Disabilities.
- *Moran, S. A. (2004). *Self-monitoring of attention and self-monitoring of performance with second grade journal writing: A comparison of two techniques*. Unpublished doctoral dissertation, University of Maryland.
- National Commission on Writing. (2003, April). *The neglected "R": The need for a writing revolution*. Retrieved April 23, 2008, from http://www.writingcommission.org/prod_downloads/writingcom/neglect-edr.pdf
- National Commission on Writing (2004, September). *Writing: A ticket to work . . . or a ticket out: A survey of business leaders*. Retrieved April 23, 2008, from http://www.writingcommission.org/prod_downloads/writingcom/writing-ticket-to-work.pdf
- National Commission on Writing (2005, July). *Writing: A powerful message from state government*. Retrieved April 23, 2008, from http://www.writingcommission.org/prod_downloads/writingcom/powerful-message-from-state.pdf
- *Nelson, J. R., Smith, D. J., & Dodd, J. M. (1992). The effects of teaching a summary skills strategy to students identified as learning disabled on their comprehension of science text. *Education and Treatment of Children, 15*, 228–243.
- *Newstrom, J., McLaughlin, T. F., & Sweeney, W. J. (1999). The effects of contingency contracting to improve the mechanics of written language with a middle school student with behavior disorders. *Child & Family Behavior Therapy, 21*, 39–48.
- Parker, R., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education, 40*, 194–204.
- Persky, H. R., Daane, M. C., & Jin, Y. (2003). *The nation's report card: Writing 2002* (NCES 2003–529). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- *Regan, K. S. (2005). *Promoting expressive writing among students with emotional and behavioral disturbance via the dialogue journal*. Unpublished doctoral dissertation, George Mason University.
- *Reid, R., & Lienemann, T. O. (2006). Self-regulated strategy development for written expression with students with attention deficit/hyperactivity disorder. *Exceptional Children, 73*, 53–68.
- *Reier, S. B. (1998). *The effects of qualitative corrective feedback on the writing skills of students with learning disabilities using electronic mail dialogue journals*. Unpublished doctoral dissertation, Utah State University.
- *Rumsey, I., & Ballard, K. D. (1985). Teaching self-management strategies for independent story writing to children with classroom behavior difficulties. *Educational Psychology, 5*, 147–157.
- *Saddler, B. (2006). Increasing story-writing ability through self-regulated strategy development: Effects on young writers with learning disabilities. *Learning Disability Quarterly, 29*, 291–305.
- *Saddler, B., Moran, S., Graham, S., & Harris, K. R. (2004). Preventing writing difficulties: The effects of planning strategy instruction on the writing performance of struggling writers. *Exceptionality, 12*, 3–17.
- Sadoski, M., Willson, V., & Norton, D. (1997). The relative contribution of research-based composition activities to writing improvement in lower and middle grades. *Research in the Teaching of English, 31*, 120–150.
- *Schmidt, J. L. (1983). *The effects of four generalization conditions on learning disabled adolescents written language performance in the regular classroom*. Unpublished doctoral dissertation, Kansas State University.
- *Schmidt, J. L., Deshler, D. D., Schumaker, J. B., & Alley, G. R. (1988). Effects of generalization instruction on the written language performance of adolescents with learning disabilities in the mainstream classroom. *Reading, Writing, and Learning Disabilities, 4*, 291–309.
- *Schumaker, J. B., Deshler, D. D., Alley, G. R., Warner, M. M., Clark, F. L., & Nolan, S. (1982). Error monitoring: A learning strategy for improving adolescent academic performance. In W. M. Cruickshank & J. W. Lerner (Eds.), *Coming of age: Vol. 3. The best of ACLD* (pp. 170–183). Syracuse, NY: Syracuse University Press.

- *Scott, K. S. (1993). *Generalization of cognitive strategies by students with learning disabilities: An instructional model*. Unpublished doctoral dissertation, University of Georgia.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22*, 221–242.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single subject research: Methodology and validation. *Remedial and Special Education, 8*, 24–33.
- Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders, 11*, 260–271.
- *Seabaugh, G. O., & Schumaker, J. B. (1994). The effects of self-regulation training on the academic productivity of secondary students with learning problems. *Journal of Behavior Education, 4*, 109–133.
- *Sexton, M., Harris, K. R., & Graham, S. (1998). Self-regulated strategy development and the writing process: Effects on essay writing and attributions. *Exceptional Children, 64*, 295–311.
- *Shimabukuro, S. M., Prater, M. A., Jenkins, A., & Edelen-Smith, P. (1999). The effects of self-monitoring of academic performance on students with learning disabilities and ADD/ADHD. *Education and Treatment of Children, 22*, 397–414.
- Smyth, J. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology, 66*, 174–184.
- *Sonntag, C. M., & McLaughlin, T. F. (1984). The effects of training students in paragraph writing. *Education & Treatment of Children, 7*, 49–59.
- *Stoddard, B., & MacArthur, C. A. (1993). A peer editor strategy: Guiding learning-disabled students in response and revision. *Research in the Teaching of English, 27*, 76–103.
- *Thanhouser, S. P. (1994). *Function over form: The relative efficacy of self-instructional strategy training alone and with procedural facilitation for adolescents with learning disabilities*. Unpublished doctoral dissertation, Johns Hopkins University.
- *Troia, G. A., Graham, S., & Harris, K. R. (1999). Teaching students with learning disabilities to mindfully plan when writing. *Exceptional Children, 65*, 235–252.
- *Van Houten, R., & McKillop, C. (1977). An extension of the effects of the performance feedback system with secondary school students. *Psychology in the Schools, 14*, 480–484.
- *Van Houten, R., Morrison, E., Jarvis, R., & McDonald, M. (1974). The effects of explicit timing and feedback on compositional response rate in elementary school children. *Journal of Applied Behavior Analysis, 7*, 547–555.
- *Van Houten, R., & Nau, P. A. (1980). The effects of two types of peer comments on the use of paragraphing in written composition by elementary schoolchildren. *Child Behavior Therapy, 2*, 55–63.
- *Wade, L. K. (1988). *An analysis of the effects of a peer feedback procedure on the writing behavior of sixth grade students*. Unpublished doctoral dissertation, University of Kansas.
- *Walker, B., Shippen, M. E., Alberto, P., Houchins, D. E., & Cihak, D. F. (2005). *Learning Disabilities Research & Practice, 20*, 175–183.
- *Wallace, G. W., & Bott, D. A. (1989). Statement-pie: A strategy to improve the paragraph writing skills of adolescents with learning disabilities. *Journal of Learning Disabilities, 22*, 541–553.
- *Weygant, A. D. (1981). *The effects of specific instructions and a lesson on the written language expression of learning disabled elementary school children*. Unpublished doctoral dissertation, University of Virginia.
- *Wolfe, L. H. (1997). *Effects of self-monitoring on the on-task behavior and written language performance of elementary students with learning disabilities*. Unpublished master's thesis, Ohio State University.
- *Zipprich, M. A. (1995). Teaching web making as a guided planning tool to improve student narrative writing. *Remedial and Special Education, 16*, 3–15.

Received October 30, 2007

Revision received March 26, 2008

Accepted March 27, 2008 ■

Appendix
Writing Instruction Treatments That Included Three or Less Studies

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Inst.	Quality Score (0-11)	PND TX	PND POST	PND MAINT	PND GEN
Dialogue journal Regan (2005)	MBD	5	SW	6 (11-12)	ED	AA (n = 2) C (n = 2) H (n = 1)	NR	Teacher	9.67	30% (Production) 20% (Quality)	35% (Production) 0% (Quality)	100% (Vocab.)	21% (Production) 38% (Quality) Setting
Direct instruction of a broad array of skills Walker et al. (2005)	MBD	3	SW	(15-16)	LD	AA (n = 2) H (n = 1)	URBAN	RES	10.00	96% ^a (Vocab.)			
McGee (1996)	MBD	5	SW	3-5	GIFTED	NR	NR	Teacher	8.67	6% (Production) 18% (Vocab.) 0% (Organize)			
Direct teaching self-regulation strategies Glomb & West (1990)	MBD	2	SW	HS	EBD	NR	NR	RES	7.17			63% (Production)	
Feedback Jerram et al. (1988)	ABAB	24	FR	5 (9-10)	—	NR	SUB	Teacher	4.40	20% (Production)			
Reier (1998)	MBD	4	SW	10-12	LD	NR	NR	Teacher	9.67	53% (Grammar)	80% (Grammar)		
Wade (1988)	ABA	17	FR	6	—	NR	RURAL	Peers	7.20	94% Group (Grammar) 50% Individual ^a (Grammar)			
Van Houten & Nau (1980)	MBD	1	SW	4	—	NR	NR	Peers	7.60	100% (Grammar)			
Goal setting for grammar / sentence construction McCurdy (2002)	MBD	9	SW	9	ADHD (n = 1) EBD (n = 1) MMR (n = 3) LD (n = 1) SLJ (n = 3)	NR	URBAN	Teacher	8.67	74% (Grammar)		38% (Grammar)	
McGee (1996)	MBD	5	SW	3 to 5	GIFTED	NR	NR	Teacher	8.67	30% (Production) 9% (Vocab.) 15% (Organize)			
Repeated writing Bauernschmidt (1991)	ABAB	3	NR	5 to 6 (10-12)	EBD and LD	NR	NR	RES	6.5	63% ^a (Production) 47% (Vocab) 27% (Grammar)			

(Appendix continues)

Appendix (continued)

Study	Design	N	Writer type	Grade (Age)	Disability	Race/ethnicity	Location	Inst.	Quality Score (0-11)	PND TX	PND POST	PND MAINT	PND GEN
Strategy instruction (revising / editing)													
Graham & MacArthur (1988)	MBD	3	SW	5 to 6 (10-11)	LD	NR	SUB	RES	8.67	64% (Total Revisions) 100% ^a (Meaning-changing Revisions)	64% (Total Revisions) 100% ^a (Meaning-changing Revisions)	84% (Total Revisions) 100% (Meaning-changing Revisions)	67% (Total Revisions) 100% (Meaning-changing Revisions) Handwriting 100% (Revisions)
Stoddard & MacArthur (1993)	MBD	6	SW	7 to 8	LD	NR	SUB	Teacher	6.50	100% (Revisions) 100% (Meaning-changing Revisions)	100% (Revisions) 100% (Meaning-changing Revisions)	100% (Revisions) 83% (Meaning-changing Revisions) Handwriting	100% (Quality) 100% (Production) Setting
Strategy instruction (other)													
Boyer (1990)	MBD	3	SW	6 (12-13)	LD	NR	NR	RES	9.58	100% ^a (Elements)	100% (Quality) 100% (Production)	100% (Quality) 100% (Production)	100% (Quality) 100% (Production) Setting
Cole (1992)	MBD	12	SW	3 to 5	LD	NR	SEMI-URBAN	RES	8.83	100% ^a (Quality) 100% ^a (Production)	100% (Quality) 100% (Production)	75% (Production) 25% (Vocab.)	100% (Quality) 100% (Production) Reading
Li (2000)	MBD	4	SW	4 to 5 (9-11)	LD	NR	SUB	RES	9.67	88% ^a (Production) 13% (Vocab.)	100% (Production)	100% (Production)	100% (Production) Reading
Nelson et al. (1992)	MBD	5	SW	4, 6, 8 (9-13)	LD	AA (n = 3) C (n = 1) H (n = 1)	URBAN	Teacher	9.67	100% (Production)	100% (Production)	100% (Production)	100% (Production) Reading
Verbal encouragement													
Kraetsch (1981)	ABAB	1	SW	(12)	LD	NR	NR	RES	7.50	100% ^a (Production)	100% (Production)	100% (Production)	100% (Production) Reading
Word processing plus													
Handley-Moore (2003)	ATD	3	SW	4 to 5 (10-11)	LD	NR	NR	RES	5.50	9% (Production)	100% (Production)	100% (Production)	100% (Production) Reading
Channon (2004)	ATD	7	SW	8 (14-15)	LD	H	URBAN	Teacher	8.83	56% (Quality)	56% (Quality)	56% (Quality)	56% (Quality)

Note. Dashes indicate that data were not reported. ATD = alternating treatment design; ABA and ABAB = withdrawal design; MBD = multiple-baseline design; N = Number of participants on single subject graph(s); AVG = average; NR = not reported; SW = struggling writer; PR = full range; ADHD = attention-deficit/hyperactivity disorder; EBD = emotional and behavioral disorder; ED = emotionally disturbed; LD = Learning Disability; MMR = mild mental retardation; SLI = speech or language impairment; AA = African American; C = Caucasian; H = Hispanic; SUB = suburban; Instr. = Instructor; RES = researcher; PND = percentage of nonoverlapping data; TX = during treatment; POST = immediately after treatment (less than 3 weeks after treatment ended); MAINT = maintenance (3 or more weeks after treatment ended); GEN = generalization (generalization to other settings or genres); Elements = basic parts of a genre or type of writing; Grammar = included number of complete and complicated sentences, sentences with adjectives, the percentage of correct placement of indentations in paragraphs, percentage of correct words in a sequence, or a more general number of correct grammar skills or points for correct writing and errors found; Production = included number of words, ideas, and/or T-units, as well as number of assignments completed; Vocab = included number of different types of words.

^a Experimental control established.