Conducting High Quality Writing Intervention Research: Twelve Recommendations

Steve E. Graham & Karen R. Harris

Arizona State University | United States

Abstract: Writing intervention research can enhance our knowledge about writing, its development, and how to teach it effectively. Despite the importance of such research, many of the writing intervention studies conducted previously were of poor quality, as documented by Graham and colleagues in a series of meta-analyses (Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007; Rogers & Graham, 2008). In this article, we offer 12 recommendations for conducting high quality intervention research, recommendations that draw on those meta-analyses as well as previous work on improving the quality of intervention research (Pressley & Harris, 1994a, 1994b) and our experiences as writing intervention researchers and editors of journals that publish intervention research (including the *Journal of Writing Research*) . The recommendations address issues and actions involved in designing, conducting, and reporting such research.

journal of WRITING RESEARCH

Graham, S. & Harris, K. R. (2014). Conducting high quality writing intervention research: Twelve recommendations. *Journal of Writing Research*, 6(2), 89-123. http://dx.doi.org/10.17239/jowr-2014.06.02.1

Contact: Steve Graham, Arizona State University, Mary Lou Fulton Teachers College, P.O. Box 872111, Tempe, AZ 85287-2111 | USA – steve.graham@asu.edu

This article presents recommendations on how to design, conduct, and report high quality writing intervention research. This is an important issue to us, as we believe that such research is a critical form of scholarship that informs both theory and practice (more about this later). This is important to us on a more personal level too, as we have devoted much of our academic careers to conducting such research (see Graham & Harris, 2009, 2012) as well as analyzing and synthesizing the writing intervention literature (e.g., Graham, Harris, & Chambers, in press).

Obviously, we love this work, and started doing it because we wanted answers. As former general and special education teachers, we were frustrated with the "pied-piper" approach to instructional practices in schools (where instructional practices are promoted and sold with passion and articulate rhetoric, but with little or no evidence of their effectiveness). We wanted to better understand how to help children become skilled writers, particularly those who found learning to write challenging (e.g., Harris & Graham, 1996).

In crafting the recommendations presented in this article, we drew heavily on our past experiences. Our meta-analyses of writing intervention research (e.g., Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007; Rogers & Graham, 2008) provided us with considerable insight into the strengths and limitations of this literature. Our experience as writing intervention researchers made us keenly aware of the many issues involved in doing this type of work well. Our collective experience as editors of five journals that publish intervention research (*Journal of Writing Research, Exceptional Children, Contemporary Educational Psychology, Archives of Scientific Psychology - Section for Educational and School Psychology,* and *Journal of Educational Psychology*) offered repeated opportunities to examine exceptional and unexceptional models of intervention studies. Our past efforts on how to improve educational intervention research broadly (Pressley, Graham, & Harris, 2005; Pressley & Harris, 1994a, 1994b) provided a platform for thinking about how to improve intervention research in writing more specifically.

Before presenting our recommendations for designing, conducting, and reporting high quality writing intervention studies, we first establish why such research is so important. It must be noted before going any further, that we mostly draw on our own research to illustrate specific points throughout this article. This is not meant to minimize the important contribution made by other writing researchers. Rather, it allows us to draw on the work we are most familiar with to illustrate our basic points and theses.

1. Why Do We Need To Conduct High Quality Writing Intervention Research?

1.1 A Compelling Practical Reason

One reason why we need more high quality writing intervention research is because writing is essential to educational, occupational, and social success, and many developing writers do not acquire needed proficiency with this complex skill. The identification of effective teaching practices through high quality writing intervention research increases the likelihood that more developing writers will become skilled writers, and can use this skill to meet school, work, and personal goals. We unpack this argument next.

Since its humble beginning as a tool to record trade goods and livestock in southern Mesopotamia over 10,000 years ago, writing is now a pervasive part of everyday life. People from all walks of life use it for many purposes. This includes using writing for communication and learning as well as artistic, political, spiritual, and self-expression. At school, writing is used to support students' learning and to assess what they know (Camps & Milan, 2000; National Council of Teachers of English, 2004). At work, writing is used by blue- and white-collar workers to accomplish job-related tasks (Boldrini & Cattaneo, 2012; Leijten, Van Waes, Shriver, & Hayes, 2014). At home, writing is used to support daily living, communication with family and friends, self-expression, and other personal goals, thanks in large part to the emergence of emailing, blogging, texting, and other forms of digital composing (Hillocks, 2006).

Despite the importance of writing, there has been a long-standing concern that many people do not develop the competence needed to use writing to meet fully and successfully the demands for writing at school and work (e.g., Sheils, 1975). This concern has its genesis in two related observations. First, many students do not develop strong writing skills. Even in an affluent country like the United States, only about one-third of students become proficient writers (National Center for Educational Statistics, 2012), professors indicate that one out of every two high school graduates are not ready for college-level writing (Achieve, 2005), and many workers do not possess the writing skills valued by their employers (National Commission on Writing, 2004). Second, writing instruction is not what it should be in most schools. Again using the United States as an example, little time is devoted to teaching writing in most schools (Applebee & Langer, 2011; Gilbert & Graham, 2010; Kiuhara, Graham, & Hawkens, 2009; Harris & Graham, 2013; Harris et al., 2012), and teachers indicate that they are not adequately prepared to teach writing (Brindle, 2012).

This set of circumstances is unfortunate, as writing is a complex skill that is not easily mastered (Rijlaarsdam et al., 2012). Skilled writing depends on the ability to regulate a variety of processes, strategies, and mental operations to plan, draft, evaluate, and revise text. The production of text involves accessing different kinds of knowledge, such as knowledge about the topic, the genre, and the intended audience. The writer must apply a host of fundamental skills, such as typing, handwriting, spelling, and

sentence construction, in order to realize their goals. Writing requires engagement and persistence, as the process is mostly self-directed. It is shaped by the writer's motivational state, as attributes such as efficacy, anxiety, and attitudes can enhance or impede how and what is created. Finally, writers must intelligently orchestrate these resources across situations, occasions, and contexts over time.

While it may be convenient to assume that good writers are born and not made, alleviating teachers and schools of the responsibility for teaching writing, the available evidence does not support this romantic proposition (Harris & Pressley, 1994). In fact, most of the practices with a proven track record for making students' better writers involve explicitly teaching students the skills, knowledge, and processes of writing (see Graham & Harris, in press).

One approach for addressing the quality gap in school writing instruction is to identify and apply evidence-based practices (EBPs; Cook, Smith, & Tankersley, 2012). These are "instructional approaches shown through high-quality research to result in generally improved student outcomes" (p. 495; Cook et al.). The primary assumption underlying this approach is that teachers should apply the best evidence available to make conscious, informed, and judicious decisions when teaching writing. This involves weighing the benefits, limitations, and even possible harm that might ensue as a result of implementing or not implementing an EPB. It also means that teachers contextualize knowledge gained from research with their own knowledge about their students, the setting in which they operate, and what they know about how to teach writing.

This approach places a premium on the development and use of proven teaching practices identified through high quality research. So why is the identification of such practices so important for addressing the practical problem of improving students' writing? The alternative to this approach is to rely on teaching lore which has not been scientifically tested.

Teaching lore is often based on practices that teachers may have experienced when they were taught to write, practices they developed and applied in their own classrooms, practices they see other teachers apply, and practices actively promoted by others as effective. This last source can come from those promoting specific approaches to writing, teachers, academics who observe teachers in action, and professional writers who draw on their own insights and experience to proffer recommendations. While individuals promoting certain approaches, teachers, those who study teachers in action, and professional writers possess considerable wisdom about the teaching of writing, basing a writing program just on such lore is a risky practice for four reasons.

First, it is difficult to separate the "wheat from the shaft" when deciding on what constitutes effective teaching lore. For example, there are many things a teacher does while teaching. As students' writing improves, they may single out one or more practices as responsible for the positive changes they observe. These may be valid choices, but it is also possible that these selections are incorrect or just correct for some students, but not for others. Academics who study teachers in action are not immune to

this selective bias, as they may overestimate the effectiveness of practices that are consistent with their philosophical views on writing and its development.

Second, writing practices based on teaching lore are often proffered with little and sometimes no direct evidence that they work (Graham & Harris, in press). When evidence of effectiveness is provided, it often takes the form of testimonials or the presentation of selected students' writing (see Smagorinski, 1987, for an example). This makes it impossible to determine if the evidence is representative or atypical.

Third, teaching lore can be based on the experience of a single teacher. In such cases, there is no way to predict if the practice will be effective for other teachers. In essence, the validity, generalizability, and reliability of practices based on the kinds of teaching lore just described are uncertain.

Basing writing instruction on practices that have a proven track record of success via scientific testing addresses the shortcomings described above. High quality intervention studies are designed to isolate the impact of a specific practice or set of practices, addressing the issue of separating the proverbial wheat from the shaft. In high quality intervention studies, evidence on whether the instructional practice produced the desired impact is systematically gathered; methodological procedures are used to rule out alternative explanations for observed effects; participants and the context in which the practice was applied are described, and statistical procedures are used to determine the confidence that can be placed in specific findings. Further, replication of effects across multiple situations is the hall mark of scientific testing. Collectively, these procedures address the issue of effectiveness, representativeness, and generalizability.

Currently, the scientific testing of writing practices using high quality research is incredibly slim. Many practices that might be effective have not been tested at all, and many other practices that have been tested have not undergone adequate replication (Graham & Harris, in press). Even for approaches that have been tested more frequently, such as teaching strategies for planning and revising, there are many gaps in the evidence-base. Just as importantly, many of the writing intervention studies conducted to date are of relatively poor quality. Thus, from a practical point of view, there is considerable need for high quality intervention research in writing.

Before moving to the next section, we would like to make it clear that writing practices based on teaching lore are not likely to disappear, nor should they. Some of today's lore will become tomorrow's EBPs, as more practices are scientifically tested. Teachers will also form judgments about the teaching lore they develop and acquire from others, and the practices that are judged as valuable will become part of their instructional repertory. As noted earlier, an EBP must be contextualized by teachers that apply them, taking into consideration their students, the context in which they operate, and what they know about teaching writing (this includes knowledge of other EBPs as well as valued teaching lore).

1.2 A Compelling Theoretical Reason

High quality intervention studies in writing do more than increase our knowledge about how to teach writing effectively. If conducted properly, such investigations can add to our theoretical knowledge about writing and writing development.

Let us illustrate this with an example. Berninger (1999) claimed that text transcription skills, such as spelling, handwriting, or typing, play an important role in early writing development. McCutchen (1988) proposed that such skills are so demanding for beginning writers, that they minimize their use of other writing processes, such as planning and revising, because they exert considerable processing demands too. Others have reported that difficulties with these transcription skills can lead children to avoid writing and develop a mind-set that they cannot write (Berninger et al., 1991). Further, devoting conscious attention to these skills when writing not only reduces cognitive resources available for other composing processes, but likely impedes their execution as well (Graham, 1990). For instance, having to switch attention while writing to a mechanical demand, such as figuring out how to spell a word, may lead a writer to forget ideas or plans held in working memory, influencing sentence construction, how much they write, and the quality of their text.

An intervention study provides an excellent vehicle for testing the theoretical proposition on whether text transcription skills play an important role in early writing development. For example, a study could address whether a program designed to improve young students' handwriting, spelling, or both, when compared to a suitable control condition, enhanced these transcriptions skills as well as students' sentence construction skills, writing output, and writing quality (see examples by Graham Harris, & Fink, 2000; Graham, Harris, & Fink-Chorzempa, 2002). If these students' transcriptions skills and other writing skills tested but not taught improved more than those of students in the control condition, this would provide support for the theoretical positions presented above. It should be noted that such a study could be conducted where students normally acquire such skills, at home or at school for example, or within a more contrived context, such as a laboratory setting, but our personal preference is that findings obtained in more naturalistic settings (everything else being equal) are more valid (see Recommendation number 5).

Instructional studies can be used to test the theoretical contribution of a variety of skills, processes, specialized knowledge, and motivation attributes to writing and its development. We provide one additional example to illustrate this idea. Models of writing for skilled as well as novice writers (Bereiter & Scardamalia, Hayes, 1996, 2011, 2012) emphasize that writing depends on the acquisition of different types of specialized knowledge, including knowledge about the intended audience, genre, and task schemas. Intervention studies can directly test this proposition by teaching such knowledge and examining if this enhances how well they write. To illustrate, Fitzgerald and Teasley (1986) taught fourth grade struggling writers about the basic elements or building blocks of a story. When compared to a suitable control group who received the same amount of instruction and writing practice, students taught this specialized

knowledge made greater gains in the quality of their compositions. For other examples applying this approach to testing theoretical propositions see Fitzgerald and Markham (1987) for revising; Harris, Graham, and Mason (2006) for planning, and Saddler and Graham (2005) for sentence construction.

Another important theoretical contribution for writing intervention research is that it can sharpen our understanding of what students' can do, expanding our conceptualization of writing development. For instance, new writing standards in the United States indicate that third grade students should be able to write an opinion that introduces the topic, states their opinion, provides reasons to support the opinion, uses linking words to connect opinion and reasons, and provides a concluding statement (Common Core State Standards, 2010). Harris et al. (2006) found that this benchmark underestimates what third grade students can do. They taught students who were experiencing difficulty learning to write a strategy for planning and drafting opinion essays and demonstrated that even weaker writers can produce opinions with all of these elements plus explanations for each reason.

In essence, we do not have a very clear idea of what students are capable of doing at different ages and grades. While there is clearly considerable variability between and within students in terms of their writing capabilities (Rijlaarsdam et al., 2012), writing intervention studies can help us better understand what students are capable of achieving under specific circumstances.

2. High Quality Writing Intervention Research

The effectiveness of a writing practice can be scientifically tested in multiple ways (Graham & Harris, in press). The gold standard of such testing is randomized control designs or true-experiments. With this type of design, students are randomly assigned to the writing treatment and the control condition. This approach provides mechanisms (e.g., randomization, control group) designed to increase the confidence that observed changes in writing performance are due to the writing treatment, as opposed to a rival explanation. The confidence that can be placed in this explanation should be viewed as a matter of degree, though, and not an either/or situation. It is possible that other circumstances not controlled by the study are responsible for the observed changes, as it would be extremely difficult to control every possible circumstance. Consequently, other approaches for testing the effectiveness of a writing practice, such as quasi-experiments (where students are not randomly assigned to treatment and control conditions), or studies where participants act as their own control (they receive both writing treatments in a counterbalanced order), can also yield a reasonable estimate that the writing treatment was responsible for observed changes in writing performance.

In addition to the designs described above, there are other methods for determining if there is a causal connection between a writing practice and observed changes in writing performance. An alternative used by some writing researchers, including us, is single-subject design methodology (Horner et al., 2005). With this approach, the

effectiveness of a writing practice is examined at the individual, not at the group level. The writing performance of each participant in the study is repeatedly assessed before as well as during and/or after instruction to establish a stable baseline and treatment effect (assuming treatment produced a positive effect). To rule out counter explanations for observed effects, the investigator controls when the writing practice is introduced to each student.

While these basic approaches for testing the effectiveness of an instructional practice differ in significant ways (e.g., random assignment, inclusion of a control group, number of participants), they share many commonalities that must be adequately addressed if a reasonable degree of confidence is to be placed in the findings from such investigations. These range from issues involved in designing a writing intervention study to conducting it to reporting it.

In the remaining pages of this article, we offer a dozen recommendations for conducting high quality intervention research in writing (see Table 1).

Table 1. Twelve recommendations for high quality intervention research in writing

1. Ask meaningful research questions
2. Test writing interventions that are well-founded and designed
3. Compare the targeted writing intervention to a credible control/comparison condition
4. Apply psychometrically sound assessments
5. Make the study as representative of the real world context as possible
6. Apply a rigorous design to answer the research questions.
7. Make certain the study is properly powered.
8. Properly analyze the data.
9. Ensure the study is conducted in an ethical manner.
10. Take steps to ensure that what is supposed to happen does happen.
11. Provide a clear, cogent, and full description of the study.
12. Design a series of studies to refine and test the writing intervention.

Our recommendations are primarily aimed at true-experiments, quasi-experiments, and single subject design studies. If a recommendation is only pertinent to a specific type of research (e.g., true-experiment), this is noted. When possible, data on how often the recommendation was violated in past writing intervention research is provided. This data is drawn from meta-analyses as previously noted. Further, some recommendations are illustrated with examples from our intervention research. The recommendations address designing, conducting, and reporting high quality intervention research.

Our decision to focus our recommendations mainly on true-experiments, quasi-experiments, and single subject design studies should in no way distract from the valuable role played by other types of research that involve writing interventions. This includes qualitative studies of the writing practices of exceptional writing teachers, case

studies examining the effectiveness of a specific writing practice, action research conducted by teachers, design research testing multiple iterations of a writing practice, interviews with teachers and students about their experiences with a particular writing method, to provide a few examples. Each of these approaches provide important information about the teaching of writing that should not be overlooked.

Recommendation 1: Ask meaningful research questions.

The purpose of intervention research is to answer one or more questions about an instructional practice. Such questions are varied and overlapping. For example, does Practice X enhance writing performance? Are observed gains from Practice X maintained over time and do they generalize to different writing tasks? Under what conditions is Practice X effective? Is practice X more effective with some types of writers than others? Is Practice X more effective than Practice Y? Are the effects of Practice X moderated by Z? What features or components of Practice X are responsible for writing gains? Do the results from teaching Practice X provide support for Theoretical Proposition A.

As an illustration, our research testing Self-Regulated Strategy Development (SRSD) for writing (Harris, 1982, 1985; Harris & Graham, 1996, 1999; Harris et al., 2009) has addressed each of the above questions at one point or another. SRSD is a multi-component intervention package that provides instruction in strategies for planning, revising, or both; teaches students the knowledge and skill needed to use such strategies as well as procedures (e.g., self-monitoring, goals-setting, self-instruction, and self-reinforcement) for regulating the use of the strategies, students' behaviors, and the writing process.

Research questions that we asked about SRSD that are similar to the more abstract ones presented above include: Is students' writing, their knowledge about writing, and their motivation to write enhanced as a result of SRSD instruction (e.g., Harris et al., 2006)? Are SRSD gains maintained over time (Graham & Harris, 1989)? Does SRSD instruction in one genre of writing transfer to another genre (Harris, Graham, & Adkins, in press)? Is SRSD effective when it is taught to the whole class, in small groups, or individually (e.g., Festas, Oliveira, Rebelo, Damião, Harris, & Graham, in press; Harris et al., 2012a, 2012b; Graham, Harris & Mason, 2005; Harris & Graham, 1985; Lane et al., 2011)? Is SRSD instruction equally effective for students with and without behavioral challenges (Lane et al., 2012)? Are the effects of SRSD instruction on students' writing moderated by students' cognitive ability (Lane et al., 2012)? Do the SRSD self-regulation components of goal setting and self-monitoring make a unique contribution to improving students' writing (Sawyer, Graham, & Harris, 1992)? Does teaching students strategies for planning enhance their overall writing performance (providing support for the theoretical role of planning in writing development; Harris et al., 2006)?

As these examples demonstrate, the questions that intervention researchers ask are typically problem focused, but they can also address theoretical issues about writing as well as teaching procedures. Intervention research questions can also be more or less meaningful. For instance, asking if sentence-combining instruction improves students' performance on a sentence-combining test is less meaningful than asking if such instruction improves students' sentence construction skills when writing. This latter query becomes even more meaningful, practically and theoretically, if the researcher also asks if such instruction results in broader improvements in students' writing (see Saddler & Graham, 2005) as well as improved reading fluency (see Graham & Hebert, 2011).

So how does an intervention researcher devise meaningful research questions? While inspiration is undoubtedly a part of this process, the best questions consistently come from those who are most knowledgeable and experienced. Meaningful questions are more likely if the researcher has a strong theoretical grounding in writing, a firm command of the available literature on this topic, a solid grasp of the literature on effective teaching and learning, and a solid understanding of schools and other contexts in which writing develops. In essence, meaningful research questions do not arise in a vacuum. Instead, they are forged through considerable study and contemplation.

Before moving to the next recommendation, it must be noted that research questions in an intervention study should be accompanied by hypothesis that are made in advance of conducting the study (a hypothesis is a statement of what will happen and why under specific conditions) . Hypotheses can be based on theory, past research, logical reasoning, or a combination of two or more of these sources. In some instances, an intervention researcher may choose not to make a specific hypothesis, as plausible arguments can be made be made for multiple outcomes. In such cases, the researcher should lay out their rationale for each of these arguments in advance and when reporting their study. Sometimes, such thinking can lead the researcher to modify their study in important ways. We illustrate this with a study conducted by Harris et al. (2005). One of our initial questions for this study was: Does teaching young writers how to plan stories and persuasive text result in improved writing for text not targeted for instruction; namely, informative and personal narrative. On the one hand, this seemed unlikely, as students were not directly taught to transfer what they were learning to the two generalization genres. On the other hand, transfer effects were possible. The instructional emphasis placed on planning across the two instructed genres might increase the value students placed on this activity, leading them to plan more frequently when writing personal narratives and informative text, resulting in better text. As a consequence of verbalizing these hypotheses in advance, we modified our study so that we had two treatment conditions that were identical in all but one way. One treatment condition included explicit instruction to promote generalization and other one did not.

Recommendation 2: Test writing interventions that are well-founded and designed.

This recommendation is equivalent to the advice given to athletes to bring your "A game." A high quality intervention study requires the investiture of considerable time, energy, and resources. While there is no guarantee that any intervention will produce the desired outcomes when tested, this is more likely to happen when the intervention is carefully and thoughtfully constructed.

Regrettably, there is no magic recipe for creating an "A game" writing intervention. There are a number of ingredients that should be part of the mix. At a minimum this should include a careful consideration of the writing task, the learner, the context, desired outcomes, instructional methods, and pertinent theoretical and empirical findings (Harris, 1982; Harris & Graham, 1996). This is not like baking a cake though, because there is no recipe for the order or amount of each of these ingredients. In fact, each of these ingredients is not really separable from the others, as they operate together when constructing an intervention.

To make this exposition more concrete, we provide an example from the second writing intervention study that we did together (Graham & Harris, 1989). We were interested in developing interventions that would facilitate writing growth for elementary-grade students who experienced unexpected difficulties with academic learning (i.e., students with learning disabilities; LD). Based on our experience as teachers working with these children (*analysis of the learner*), we had noted that their oral stories were much more expansive and complete than their written stories. When writing stories, they produced very little text, gave little to no forethought to the process of writing, created incomplete stories, and commented negatively about their writing capabilities. They did not appear to put much effort into the act of writing in most of the instances we observed.

In thinking about our experiences in working with these students, we knew that there were many possible reasons for what we had observed firsthand. For instance, students produced little text and put little effort into what they wrote because they did not value writing or the types of writing they did at school. Likewise, slow handwriting and frequent misspelling, a common issue for many of these children, impeded the story writing process.

We focused our attention in this particular study, however, on a different set of culprits. This was driven by two different, but compatible theoretical models of writing and writing development (*analysis of pertinent literature and theory*). First, a few years before we designed our study, Hayes and Flowers (1980) had specified the mental operations skilled writers apply as they compose, indicating that they developed plans for what to say and how to say it, and that their planning involved setting goals, accessing possible writing content from memory, and deciding how to organize it. Second, Scardamalia and Bereiter (1986) proposed that novice writers apply a greatly simplified version of this process, simply converting writing into telling what one knows. Their search for relevant content from memory is driven by what they know

about the topic and the type of text to be produced (e.g., genre knowledge). This second description was consistent with our observations of how students with LD typically write. As a result, we reasoned that these children did not skillfully access the relevant knowledge they possessed when writing a story, and that one possible means for helping them do so was by teaching them a strategy for planning what to say and how to say it (something done by more skilled writers).

To devise this planning strategy, we examined the empirical literature on story writing (*analysis of the task*). In the late 1970s, research conducted by Stein and Glenn (1979) showed that the grammar or structure of stories can be reduced to two major components: setting and episode. A setting commonly involves three elements: introduction of the main character(s), information about the time of the story, and description of the locale. An episode includes a starting event (a precipitating event occurs), goal (the main character(s) reacts to the starting event and formulates a goal), action (planned efforts to achieve the goal), ending (result of the action), and reaction (character responses to actions or the consequences). Using this information, we designed a planning strategy where the student is asked to generate possible ideas for their story prior to writing by identifying the main character(s), when and where the story took place, what the main character wanted to do, how the story ended, and how characters reacted. Thus, students with LD would use genre knowledge about stories to help them plan what to say and how to say it. This strategy took advantage of their existing knowledge as well as knowledge they would develop as a result of instruction (see below).

As we began thinking about how to teach this planning strategy to students with LD, we had to consider again what we knew about them (*analysis of the learner*) and how to make sure they learned to use the strategy successfully (*analysis of instructional methods*). We provide three examples here to illustrate this. One, to use the planning strategy successfully, students needed to be familiar with the basic elements of a story. This was accomplished by having them read stories and identify these basic elements until they could generate examples for each element. Two, because we were asking students to replace their current approach to writing with an approach that required more effort, we put into place mechanisms to make the value of the planning strategy evident to them and to increase their sense of efficacy. The students in our study were asked to evaluate improvements in their story writing and were encouraged to attribute these improvements to effort and use of the strategy. Three, children with LD benefit from instruction that scaffolds and supports their learning (Swanson, Harris, & Graham, 2013). As a result, instruction in how to use the planning strategy involved a gradual release model, where the instructor initially explained the rationale behind the planning strategy, modeled how to use it, and worked conjointly with students to apply it, with the instructor gradually withdrawing support until students could apply the strategy independently and effectively.

Throughout this whole process, we constantly returned to what we viewed as the ultimate goal of our intervention: to improve the story writing of students with LD

(*analysis of outcomes*). As noted earlier, these children's stories are typically brief and incomplete. They do little planning in advance, and they are often negative about their capabilities as writers. Our goal was to address each of these issues. Thus, we decided that it was critical to assess if our writing intervention resulted in longer, more complete, and qualitatively better stories immediately after instruction and over time. We also decided to assess if there was a positive change in students' self-efficacy for writing as a result of our intervention.

Interventions are delivered in specific contexts (*analysis of context*). Each context has its own unique attributes, and it is important to take these into account when designing an intervention. For example, we selected a task, story writing, that was common and valued in the schools where the study took place. Writing instruction for students with LD in these schools mostly took place in small groups, so the instruction for this study was designed with that in mind. No two instructors are identical. As a result, the lessons for the writing intervention were presented as a meta-script. Instructors were asked to follow the basic script and complete all specified activities, but were free to make needed modifications to an activity to ensure it was successful for their students. This might include additional explanation, modeling, feedback, and so forth.

The description above represents a linear and simplified version of a recursive and more complex process we undertook in designing the writing intervention for Graham and Harris (1989). Even so, it provides a window into the ingredients needed to design "A game" writing instruction. We would like to highlight an additional activity that is important in designing such an intervention: a theory of change that shows, at a minimum, the linkage between learner, instruction, and outcome. For example, the theory of change for writing self-efficacy in the study described here was that students with LD are not positive about their story writing capabilities (learner) and teaching them strategies for successfully writing stories and making their progress concrete through self-monitoring (instruction) would result in a stronger sense of efficacy (outcome).

Of course, not all writing practices are designed by the research team that scientifically tests them. Nevertheless, a careful analysis of the writing intervention to be studied in terms of task, learner, context, outcomes, methods, and pertinent theoretical and empirical findings is still recommended, as such analyses will result in some needed adaptations to the treatment.

Lastly, there are many different writing interventions and potential writing interventions that can be tested scientifically (Pressley et al., 2006). Such interventions vary in terms of complexity, ranging from short focused interventions to complex, multicomponent treatments carried out over a series of years. The scientific testing of broad, multi-component writing interventions across grades is rare (see Graham et al., 2012; Graham & Perin, 2007), but such tests are needed, as the potential impact of such treatments are considerable. Even so, much can be gained by testing more focused writing treatments, as this research provides needed information for constructing more

encompassing writing programs and, in some cases, informing theory and development (see above).

Recommendation 3: Compare the targeted writing intervention to a credible control/comparison condition.

While this recommendation does not apply to all types of writing intervention research (e.g., single subject design studies), it is especially important when designing true- or quasi-experiments. The outcomes in such studies depend upon what the treatment is compared against. For example, a more positive outcome is likely when a writing treatment is compared against no instruction or a "straw man" condition (i.e., poorly operationalized control/comparison condition). In contrast, a researcher is less likely to obtain a positive outcome when a more credible comparison condition is applied; that is, one that is as similar as possible to the writing treatment with respect to variables that do not differentiate these conditions by definition (Pressley & Harris, 1994).

Unfortunately, there are too many writing studies employing weak comparison conditions. For instance, Graham and Perin (2007) reported that 43% of true- and quasi-experiments testing the effectiveness of the process approach to writing instruction involved a no instruction control condition or a poorly specified comparison condition. Even when comparison conditions are better described, they often involve a weak comparison. As Pressley and Harris (1994a) noted, researchers commonly create the best possible version of their preferred treatment (as suggested in Recommendation 1), but fail to devote similar efforts when creating comparison conditions. In some instances, they may not understand the comparison condition well enough to construct a viable instantiation of it. In other instances, they may create a comparison condition with confounds that undermine the interpretability of the study. To illustrate, the writing treatment and control conditions may differ in terms of amount of instructional time, type of writing students are assigned, or instructors' prior experience with one or the other conditions. Confounds such as these, make it difficult to determine if improved writing outcomes are due to the tested writing intervention or the confounding variables.

As a result, the best course of action is to design a comparison condition that is as similar as possible to the tested writing intervention in terms of the variables that do not differentiate the two conditions by definition. As an example, we conducted a study testing handwriting instruction (Graham, Harris, & Fink, 2000), where we carefully crafted the comparison condition, so that it allowed us to address out major theory-driven question: Does handwriting instruction improve how much young children write, their sentence construction skills, and the quality of their writing? To answer this question, we needed a comparison condition that did not involve handwriting instruction and was unlikely to enhance handwriting performance, but was comparable with regards to type of instruction, duration of instruction, student s' level of motivation, and experience of instructors. Our comparison group involved one-on-one

explicit instruction in phonological awareness for 15 minutes a day for 27 days delivered by a trained tutor. With one exception, the teaching of handwriting, these same conditions were in effect in the treatment group. One benefit of this approach was that all participating students received instruction designed to improve some aspect of their literacy skills. Another benefit was that all students and instructors were told they were participating in a special program, providing a possible control for Hawthorne effects (i.e., children may work harder when they believe they are part of an experiment).

Before moving to the next recommendation, we would like to comment on the use of "business-as-usual" comparison groups. While such control groups have obvious limitations, they are not likely to disappear. They provide a relatively cheap alternative to constructing a well-designed comparison condition. Moreover, many schools do not allow researchers to implement multiple interventions in their classrooms (at least in the United States). When such a comparison condition is used, it is unlikely that all students in the comparison classrooms are receiving exactly the same instruction. A good rule of thumb for business-as-usual comparisons is to collect as much information as possible about what happens in all of these classrooms, providing a "better chance of explaining why interventions work relative to comparison conditions (when they do) and why they fail (when they do not work)" (Pressley et al., 2006, p. 6).

Recommendation 4: Apply psychometrically sound assessments.

The best laid plans for an intervention study can come to naught if the measures used to assess the outcome of the treatment are flawed. At the most basic level, measures must be reliable. Given that some writing measures (e.g., writing quality) are somewhat subjective and others are quite complicated (e.g., mapping the organizational structure in a composition), we find it disquieting that close to 30% of studies in one meta-analysis that Graham conducted with a former student (Sandmel & Graham, 2011) did not establish scoring reliability for measures of writing quality. More concerning is the observation, based on analyzing writing intervention studies in close to 15 meta-analyses (see Graham & Harris, in press), that researchers rarely seek to establish other forms of reliability for their dependent measures (e.g., coefficient Alpha, test-retest reliability, or alternate form reliability).

While reliability is important, a reliable measure is not necessarily a valid one. Outcomes must be a valid measure of the construct they are designed to measure if the impact of a writing treatment is to be credibly evaluated. In analyzing thousands of writing intervention studies, we observed that this is established much less often than reliability. Consequently, an important aspect of designing a high quality writing intervention study is making sure that measures are both valid and reliable (see van Steendam, Tillema, Rijlaarsdam, & van den Bergh, 2012). If measures are not psychometrically sound, then statistical power declines and the internal and external validity of the investigation collapses (Levin, 2005).

In selecting or developing measures for a writing intervention, researchers need to consider issues involving floor and ceiling problems (i.e., students' scores cluster at the lower or upper end of the test). While only about 10% of studies in Graham and Perin (2007) evidenced floor/ceiling problems, close to 25% of studies in Graham et al (2012) exhibited this problem. As Levin (1985) indicated, ceiling/floor problems can prevent the researcher from precisely determining the impact of the target intervention, and in some cases make it impossible to test the treatment at all (e.g., measures are already at ceiling levels before instruction is initiated). With ceiling/floor problems, the test can be too easy, too hard, or contain too few items at one end of the test or the other. One partial solution to this problem is to select measures where such issues are unlikely. A second is to pilot potential measures with students who are similar to the ones participating in the planned study.

Floor effects may be especially problematic for pretests assessing behaviors that students engage in rarely. For example, some groups of students such as children with special needs or students writing in a second language may limit their revising to changes involving single words and phrases, making few or no revisions involving larger units of text (Van Steendam, Rijlaarsdam, Van den Bergh, and Sercu, in press; MacArthur & Graham, 1987). Situations like these put a writing researcher in a tough position. They can decide not to pretest students on this variable (assuming that students' performance is low and skewed). They can attempt to develop an alternative measure designed to eliminate floor effects (in some cases this may be very difficult to do). Or they can use the pretest, acknowledging that it is flawed and tempering their conclusions accordingly.

In deciding what to assess in a writing intervention study, a good rule of thumb is to assess what was taught as well as more general measures of writing performance. For instance, if students are taught specific spelling words and skills during the course of a study, the researcher should assess if these skills improved from pretest to posttest. It is also a good idea to assess whether there were more general improvements in spelling. Depending upon the purpose of the study, assessing improvements in reading and writing may be advisable too (e.g., Graham et al., 2002).

Likewise, if students are taught how to plan a story by generating possible writing ideas by considering each basic element of a story in advance of writing (Graham et al., 2005), the investigator may want to determine if these elements become more evident in students' writing after treatment than they were before it, and more broadly assess if (1) the quality of students' story writing improves; (2) the quality of their personal narratives improve (a close but not directly taught genre); (3) knowledge about the basic parts of a story expands; and (4) self-efficacy about story writing increases.

In addition, students' performance on the most critical outcome measures should be assessed over time to ascertain if treatment effects are maintained. With the exception of strategy instructional writing research (see Graham & Harris, 2003; Graham, Harris, & McKeown, 2013), writing intervention researchers rarely assess maintenance effects (Graham et al., 2012; Graham & Perin, 2007).

The inclusion of pretests, posttests, and maintenance assessments introduces another issue that should be considered when planning an intervention study. The measure for a particular construct should be equivalent at each assessment point. For example, if students are asked to write an opinion essay about a particular topic after reading source material before the intervention, after it, and six months later, then three equivalent assessments for this skill are needed. In designing such prompts for students, the researcher will want to make sure that the three prompts are as similar as possible (except for the topic), and students similar to the target participants view writing the topics as interesting, can read and understand the source material, and write a reasonable opinion essay after reading it. Beyond such initial development and field testing, the researcher should further establish empirically that the three prompts produce similar levels of performance. Even with all of these safeguards in place, the safest course of action is to counterbalance the administration of the three assessments, so that one-third of students is administered each assessment at each time point (pretest, posttest, and maintenance).

An equally important issue is how many writing prompts should be administered at any given time or assessment point? It does not appear that a single piece of writing provides a reliable estimate of overall writing achievement (Graham, Harris, & Hebert, 2011a). This is recognized in single subject design research where a minimum of three pieces of writing are collected during baseline and treatment. Consequently, the safest course of action is to collect multiple writing samples at each testing point.

If at all possible, the persons administering assessments should not be the same person delivering instruction, as they have a stake in the outcome of the study. When scoring writing samples, it is equally important that all identifying information be removed and that those scoring the assessments are not familiar with the participants or the purpose and design of the study. Depending on the purpose of the assessment, the researcher may want to have all compositions typed and mechanical errors corrected. In a meta-analysis by Graham, Harris and Hebert (2011b), differences in legibility and number of spelling miscues in a paper can raise or lower scores on assessments of the quality of the content in a paper by a full or one-half a standard deviation, respectively. Lastly, when scoring students' writing, papers should be randomly ordered, as judgments on the quality of a paper can rise or fall depending on the quality of the papers that were scored just before it (Graham, Harris, & Hebert, 2011a).

Recommendation 5: Make the study as representative of the real world context as possible.

The credibility of a writing intervention study is enhanced if the research takes place under normal conditions (i.e., in schools and other venues where writing develops). This demonstrates to teachers, policy makers, and others that the findings from a study are relevant to real world situations. While a researcher may decide not to conduct the first test of a writing intervention under such representative conditions (see

Recommendation 12), the ultimate goal of intervention research is to see if a writing practice is effective in typical learning situations.

Before moving forward, we would like to make two points. One, teaching students a writing skill they are not likely to use at that immediate point can be a questionable decision, but there may be situations where doing so can be important. For example, we taught second and third grade children how to plan and write opinion essays (Graham et al., 2005; Harris et al., 2006) at a time when primary grade students were rarely give such assignments at school (Cutler& Graham, 2008). However, we wanted to determine if younger students, particularly ones who found writing challenging, could learn to write such compositions, as we felt that the writing curriculum at these grades was too narrow.

Two, representativeness is enhanced if participants and implementation contexts are randomly sampled or selected so that they are similar to the populations/contexts to which generalization is desired (Levin, 2005). Whenever possible, we encourage writing intervention researchers to employ such strategies.

Recommendation 6: Apply a rigorous design to answer the research questions.

Too often, writing intervention researchers apply a weak design to answer their research questions. For example, in two meta-analyses of true- and quasi-experiments in writing (Graham et al., 2012; Graham & Perin, 2007), Graham and colleagues found that randomization occurred in just about one-third of all studies; teacher effects were only addressed in about two out of every five studies; and more than one third of studies included just one class per condition. While writing interventions tested via single subject designs were more rigorous, they were not impervious to design problems, as about 10% of all studies did not collect an adequate number of writing probes before, during, or after treatment (Rogers & Graham, 2008).

Issues such as these undermine the internal validity of a study, decreasing the confidence that can be placed in causal claims that the outcomes of a study are attributable to the writing intervention. As a result, it is important to apply the most rigorous design possible when conducting an intervention study. For group studies, where some students participate in a treatment condition and other students participate in one or more control/comparison conditions, it is best if students are randomly assigned. As Levin (2005) indicated, this "is an absolute necessity to ensure initial equivalence of groups" (p. 16). If the unit of assignment is larger than the individual (e.g., class, school), randomization at the level of assignment is still preferred.

If randomization is not feasible, it may be possible to match carefully students, classes, or schools carefully with respect to relevant variables (e.g., writing achievement, SES, gender) so that those receiving the writing treatment condition can be compared to those who are similar, but did not receive it (Pressley et al., 2006). At the very least, participants should be pre-tested prior to implementing the writing

intervention and control conditions to determine if they are equivalent on relevant variables. Graham and Perin (2007) found that this occurred in only 57% of the studies in their meta-analysis. It should also be noted that using overall academic achievement or grades at pretest as proxy measures for writing achievement is risky, as such proxies may show little relationship to posttest measures of writing (Braaksma, Rijlaarsdam, & Van den Bergh, 2014).

While quasi-experimental evaluations are likely biased, as it is virtually impossible to establish initial equivalence on all relevant variables in such investigations, quasi-experiments can still be of value if equivalence is established for the most relevant variables, as a number of meta-analyses have shown that true- and quasi-experiments yield the same effects much of the time, with quasi-experiments producing slightly more variable effects (e.g., Lipsey & Wilson, 1993; Larzelere, Kuhn, & Johnson, 2004).

Even when randomization does occur, pretesting participants on relevant variables to verify equivalence of the writing treatment and control/comparison condition(s) is recommended when the sample size for a study is small. Random assignment does not guarantee that groups are equivalent. It only ensures that differences are due to chance. This assumption is less likely to be valid with smaller samples. The value of pretesting in true-experiments is less certain, however, for intervention studies of short duration. Pretests can influence students' performance on subsequent writing posttests (Cambell & Stanley, 1963), and this is more likely to happen when there is little time between the testing points.

In addition to putting into place procedures designed to increase the likelihood that preexisting characteristics of the groups compared are equivalent, intervention researchers should also identify and address other confounds that cloud interpretation of the results. The basic rule of thumb is to minimize or control as many confounds as possible. Possible confounds that can influence the outcome of a study, weakening claims the treatment was responsible for improvements in writing performance include: difference between treatment and control/comparison conditions in (1) who delivers instruction, (2) amount of time spent receiving instruction, and (3) types of writing activities assigned (e.g., one group learns to plan persuasive text and is compared to a control group that writes stories).

Other confounds involve participants' expectations. If teachers or students in the writing treatment group believe they are delivering or receiving special instruction (i.e., Hawthorne effect), they may work harder than those in a control group who are not made to feel special. In contrast, teachers or students who are aware that they are in a control condition and are being compared to those in a special treatment, may work harder to overcome the disadvantages of being a control participant (i.e, John Henry effect). Still other possible confounds include possible variations or drift in terms of how the writing treatment or control/comparisons conditions are implemented as well as contamination (i.e., teachers in the control condition start applying procedures used by treatment teachers).

In creating the strongest design possible to answer research questions, there are a number of strategies that intervention researchers can apply. They can hold potentially confounding variables constant (e.g., instructors teach students in each condition, all groups receive the same amount of instruction; all groups are told they are part of a special instructional group). They can also make potentially confounding variables part of their design (e.g., both students and instructors are randomly assigned to conditions and possible teacher effects are tested statistically). Further, they can construct protocols for participants and research staff, specifying what each does. Such protocols can include procedures for controlling for instructional drift, contamination, and expectations. Of course these strategies are only useful, if potential confounds that need to be minimized or controlled are first identified.

Recommendation 7: Make certain the study is properly powered.

Twenty years ago, Pressley and Harris (1994a) drolly commented, "There is one way to be certain of obtaining statistically significant differences: Run studies with a large number of subjects, so that the power for detecting even miniscule effects is extremely high" (p.199). Of course, the opposite problem exists. A study with a very small number of participants is unlikely to obtain statistical significance unless the effect of the writing treatment is very large.

Given this state of affairs, it is surprising that writing intervention researchers rarely conduct a power analysis before conducting their study. This involves using statistical procedures to determine the minimal number of participants needed to detect an experimental effect of a specific size. In all of the writing intervention studies we analyzed over the year, less than one in thirty investigations reported using such a procedure. Given the cost and time involved in conducting a writing intervention study and the possible rewards for obtaining statistically significant effects (i.e., publication), this is a perplexing paradox, as a relatively simple procedure makes it more likely that a researcher adequately tests their preferred writing intervention. Consequently, investigators should determine in advance what counts as a credible difference or effect size and determine the number of participants needed so that the study is adequately powered.

One problem that can weaken the power of a study to detect a meaningful difference between treatment and control/comparison conditions is attrition (i.e., the loss of participants in a research study). Graham et al. (2012) and Graham and Perin (2007) reported that 20% to 30% of studies in their meta-analyses of true- and quasi-experiments in writing evidenced attrition above 10%. If attrition rates are high enough (20% or greater), they can change the statistical outcome for a study (Stinner & Tennet, 2012) and cloud external validity. As a result, means for reducing attrition should be considered when planning an intervention study. For example, gauging teachers' commitment to participate in a study, regardless of the condition they are assigned,

may reduce teacher attrition. Providing a monetary incentive for participation may also increase the probability that teachers will see the study through to its end.

Further, intervention researchers need to specify explicitly who will participate in the study, and develop a plan for collecting relevant information about them. For students, this includes data on writing achievement, gender, socio-economic status, grade (age), race, and so forth. For teachers, it includes previous training, education, experience, gender, race, and other relevant variables. Collecting such information is important for two reasons: (1) establishing internal validity (i.e., participants in each condition should be comparable) and (2) external validity (i.e., a rich description of participants makes it possible to consider more accurately who the results generalize to).

Recommendation 8: Properly analyze the data.

There are many different statistical procedures for analyzing data from any single investigation. Too often, however, data is not properly analyzed in writing intervention studies. To illustrate, in two meta-analyses of writing intervention research involving true- and quasi-experiments with students (Graham et al., in press; Graham & Perin, 2007), we examined if the proper unit of analysis was used in the statistical analyses. A unit is the smallest entity to which a treatment is assigned. It can involve a student, a small group, a class, or school that receives instruction. This is important because virtually all ANOVA-based statistics are based on the assumption that units are independent (Levin, 1985). Thus, if classes are assigned to treatment and control conditions, the proper unit of analysis is the mean of the class, as one student's performance in a class is dependent, at least in part, on other students' performance (they were taught and likely tested together). While the reports of these two meta-analyses did not indicate how often the unit of analysis assumption was violated when classes or schools were assigned to conditions, this assumption was violated in almost every case.

In addition, few researchers tested the assumptions underlying the statistical procedures employed. The application of ANOVA-based statistics, which were used in the large majority of these studies, depends on a number of assumptions being met, including normal distribution and equivalent variability for the dependent variables in each experimental condition (Levin, 1985). With ANCOVA, which was applied in a sizable minority of the investigations, it is further assumed that the slopes for the interaction of covariate and each respective experimental condition are equivalent (i.e., homogeneity of regression slopes assumption). If these assumptions are not met, the resulting analyses can lead to erroneous conclusions.

Much of the writing intervention research conducted to date involves individual students nested with classrooms or schools. New statistical tools that take into account the nested nature of such data are now available (Bryk & Raudenbush, 2002). Such multilevel models address issues involving the appropriate unit of analysis, and they

can allow researchers to examine individual differences in growth if performance is repeatedly measured across time. These procedures also provide an alternative to ANCOVA, without the requirement of meeting the homogeneity of regression slopes assumption. As a result, these statistical procedures need to become ubiquitous in writing intervention research, as they address many of the concerns presented above.

In developing a high quality writing intervention study, the researcher should plan their analyses before data are collected and, as Pressley and Harris (1994a) noted, they need to stick "with the plan unless there is a very good reason to deviate from it" (p. 198). For example, the plan may need to be modified if data are not normally distributed or considerable heterogeneity across conditions is evident (the latter is not a major concern if sample size for the different groups are the same or close to the same). By developing an a priori versus a post hoc plan for statistical analyses, there is less likelihood that researchers apply multiple statistical schemes in search of the one that provides evidence that their preferred writing treatment works.

Of course, the planned statistical procedures must align and be appropriate for the research questions and design applied by the researcher (just as the research design must align and provide a suitable vehicle for answering the proposed research questions). Fortunately, there are many suitable procedures to choose from, including t-tests, ANOVA, ANCOVA, regression, and multi-level modeling (e.g., HLM and growth curve analysis). No matter what method is chosen, the assumptions underlying the use of the statistical procedure should be tested and adjustments made to the significance level of the test if there are more than two experimental conditions (e.g., a Bonferroni adjustment).

Recommendation 9: Ensure the study is conducted in an ethical manner.

Instead of concentrating on frequently discussed ethical responsibilities in conducting research (e.g., obtaining consent and assent, maintaining confidentiality, storing data properly, keeping meticulous record of study activities, ensuring integrity of data, and determining authorship), we would like to address one issue in particular. In writing intervention research, participants spend valuable time engaged in the activities of the study. As a result, researchers need to think carefully about what they ask participants to do. For example, engaging control/comparison students' in an intervention that is purposefully designed to be weak (i.e., straw man intervention) is not defensible. As Pressley and Harris (1994a) observed, "there are too many [such] scarecrows habiting the pages of journals publishing educational intervention research" (p. 196).

Consideration should further be given to whether the writing intervention is worthy of the time devoted to it, and whether the amount of time spent testing students is defensible, especially if it takes participants away from valuable instructional time. Moreover, researchers may want to consider applying a cross-over design where students in both the treatment and the control condition receive the writing intervention (cf. Harris et al., 2012a).

Recommendation 10: Take steps to ensure that what is supposed to happen does happen.

Conducting high quality research depends upon attention to detail. Once the study is in progress, the researcher must pay close attention to what happens during the writing intervention, the control condition, and testing. This requires putting into place procedures for ensuring correct implementation as well as collecting evidence to demonstrate (or not) that such implementation did occur.

In terms of developing implementation procedures, protocols for delivering the intervention, control/comparison conditions, and assessments should be developed and evaluated before the start of the study. It is important to try them out to make sure important steps are not missing. For the writing intervention and any designed control/comparison conditions, the protocol typically involves one or more lessons that include the pertinent procedures and materials for carrying out the intended instructional activities. Similarly, protocols for administering assessments provide directions for how and in what order to carry out the planned assessments (this often includes directions for what to do to establish rapport as well as what to do when things do not go according to plan).

Just developing such protocols is not enough, however, as teachers, instructors, and raters need to practice implementing intervention procedures and components until they can do them correctly and independently. This involves several steps: (1) establishing criteria for demonstrating mastery of each protocol; (2) developing needed training materials; and (3) and practice applying each protocol until mastery is achieved. It is important that those carrying out the study can apply these protocols correctly and in a natural manner.

Views differ strongly, however, as to how intervention protocols should be delivered, with proponents differing as to whether a scripted, standard protocol approach or a problem solving approach will be more effective, although no research has been done comparing the two approaches. The standard protocol approach requires the use of a fully scripted, standard protocol for a preset period of time when an intervention is implemented by teachers or others. Teachers read directly from the script, cannot teach for less or more time than indicated, and are not allowed to make changes or differentiate instruction beyond what is in the script. The problem-solving protocol approach, on the other hand, allows teachers and others to implement interventions with flexibility to differentiate to meet individual needs and without being expected to read from a set protocol or teach for a set period of time. In our SRSD research, we have always used a problem solving protocol approach rather than a scripted protocol approach, which has been successful.

It is unclear how often or to what degree project staff received training to implement study procedures in past writing intervention research. Graham and Perin (2007) noted that less than one-half of investigators reported providing training to teachers/instructors. While this may be accurate, it is also possible that such training did occur, but was not reported by researchers.

Somewhat similarly, investigators typically do not provide data that their treatment was delivered as intended. For true- and quasi-experiments this was done in less than 30% of past studies (Graham et al., 2012; Graham & Perin, 2007). It occurred at a slightly higher rate (43%) for single subject design studies (Rogers & Graham 2008). Such data is critical in determining if the writing intervention was in fact responsible for improved writing. A basic assumption underlying intervention research is that the intervention was carried out as intended. If this is not established, then the confidence that can be placed in treatment effects is greatly diminished.

There are a number of approaches for establishing treatment fidelity. Teachers/instructors can be asked to check off each activity as they complete them or keep a log of what they did each day. Even better, instructional sessions are observed, audio-recorded, or video-taped and an independent observer record (often with the help of a checklist) what occurred during instruction. This can include whether each intended activity occurred as intended as well as the quality of instruction delivered. Of course, more confidence can be placed in these observations if reliability is established. While there is no established criterion for what percent of lessons should be part of the processes for establishing treatment fidelity, we typically collect data on one-fourth to one-third of all lessons.

For true- and quasi-experiments, treatment students are compared to students in a control/comparison condition. Rarely, in our experience, do investigators document what happens to students in this latter condition. It is important to do so for three different reasons. One, it is difficult to gauge the effectiveness of a treatment without knowing what it is being compared against (as noted earlier). Two, without documentation, the possibility of contamination is not addressed, as control teachers may apply one or more features of the treatment in their classroom. Three, the opportunity to detect possible confounds that arise during the course of the study is diminished, as such confounds are not limited to the treatment condition.

Recommendation 11: Provide a clear, cogent, and full description of the study.

All of the brilliance, hard work, and resources that are put into a high quality writing intervention study yields few or no returns if the investigation is not made available to the larger research community and public or if it is so poorly written that it is difficult for others to understand. In conducting various meta-analyses in the arena of writing, Graham was stunned by how few dissertations or thesis were published as journal articles (even though the quality of these works was often equal to published studies and sometimes exceeded them).

We would like to offer a few recommendations with regards to producing a writing intervention research paper that is publishable. It is important to realize that the researcher needs to make an argument and tell a story. The argument begins in the Introduction where the researcher establishes the purpose of the study (questions and

predictions) and why the investigation and each of its components are important. The story also starts here as the researcher needs to set the study within the context of the previous literature (e.g., telling how the study replicates prior research, extends it, or both) and tell how the study and the questions it answers are theoretically grounded. The argument and story must be lean and to the point, as neither reviewers nor readers have much tolerance for extraneous information or details.

The story continues in the Methods section of the report, where the researcher reveals who was involved in the study, describes the setting in which it took place, details what happened in the treatment and control conditions, explains how students were assessed, and provides a chronology events. It is essential that the story is rich and deep enough so that it can be understood and replicated (Levin, 2005). For studies involving complex interventions, a table showing how the writing treatment(s) and control/comparison conditions are similar and different makes it easier to follow the story. In some instances, it may be necessary to make fuller descriptions of the treatment as well as researcher designed assessments available in other venues to ensure replicability (some journals provide a space on their web site where such materials can be stored and accessed). It is absolutely essential that the story told includes information about the reliability and validity of each measure. It is further important to keep in mind that reliability is a characteristic of the sample and not the measure (Crocker & Algina, 1986).

The meat of story is in the Results section, as this is where the researcher tells what was discovered. The story is often easiest to understand if the findings are presented for each research question separately. At this point, the story needs to specify how each question was analyzed statistically, what was discovered, and whether or not the finding was consistent or inconsistent with the predictions established in the first section of the paper. Just as importantly, appropriate statistics need to be reported. This includes the results of the statistical analyses as well as other basic outcomes such as the mean, standard deviation, confidence interval, and effect size for each measure. As indicated earlier, commonly omitted from this story is whether the assumptions underlying the statistical analyses were met and, if not, how they were addressed. The story is not complete without this information.

The argument is resumed in the final section of the paper where the findings of the study are dissected and discussed. In essence, the researcher addresses whether the predictions made at the start of the report are supported by the findings from the study. When predictions are not supported, the argument typically shifts to consider what are the most likely explanations for the unexpected findings? The story is also resumed, as the findings are situated within the pertinent literature, recommendations for future research are drawn, and implications for practice are considered. A caution is in order, though, as what one is entitled to conclude should *actually* match what one does conclude.

Recommendation 12: Design a series of studies to refine and test the writing intervention.

When Graham and Perin (2007) finished the first broad review of the writing intervention literature, they were surprised at how seldom a researcher or research team sought to replicate their findings for a specific writing practice. This was also the case in subsequent reviews (Graham et al., 2012; Rogers & Graham, 2008). While there were some exceptions, most of the accumulated studies involved a single researcher or research team conducting a single intervention study, leaving any additional testing to others.

While greater confidence can be placed in the effectiveness of a writing intervention when its impact has been confirmed by multiple research teams under different conditions, it is also desirable that the same research team tests its effectiveness more than once to demonstrate that it is worthy of additional study. Further, when designing a writing intervention, it is important to test and refine it, just as engineers do when developing a solution to a problem.

We try to apply this logic in our writing intervention research (see for example Graham & Harris, 2005; Graham, Harris, & Zito, 2005). As we initially design a writing intervention, we may test one or more parts of it with a few children to make sure that it is age-appropriate and works. If not, we redesign those parts. Once the intervention is fully assembled, we generally test it and refine it through one or more design experiments or a series of single subject design studies. With design experiments, we apply the intervention and modify it as needed based on our observations as well as those of the teachers and participating students. We further collect relevant outcome data before and after the application of the intervention to get a rough gauge of its effectiveness. With single subject design studies, we try to conduct at least two separate studies, modifying the treatment after each investigation. These refinements are again based on our evaluation and teacher and students feedback. An advantage of single subject design studies is that they provide credible evidence that the treatment was responsible for changes in students' writing performance (see Horner et al., 2005).

Once we have refined the writing treatment, we then test it on a larger scale applying a true- or quasi-experimental design. We often conduct the first study with our own graduate students delivering the intervention to be sure that it works in this broader context (this usually occurs across different schools). Next, we test it with teachers delivering the intervention. While it is not always possible, we like to follow such studies up with component analyses to determine if specific features of the treatment make a unique contribution to improving writing outcomes (see for example Sawyer, Graham, & Harris, 1992; Harris et al., 2006).

We would also like to note that we commonly collect qualitative and quantitative data in our *intervention* studies. The qualitative data involves interviews, observations, or both with teachers and students during the study to determine what participants thought was happening during the intervention, whether they viewed it as valuable, and what they viewed as its strengths and limitations, including suggestions for

modifying it (Harris, Lane, et al., 2012; MacArthur, Schwartz, Graham, Molloy, & Harris, 1994 ; McKeown, 2012).

3. Concluding Comments

Research, especially intervention research, is not a simple process. It involves a host of interrelated decisions and actions involving purpose, participants, setting, treatment and control/comparison conditions, assessments, design, statistical methods, and ethical issues. Despite how difficult it is to conduct intervention research, it can be done well and it is essential to the study of writing. It provides critical information about writing, how it develops, and how to teach it.

The present article provides 12 recommendations for conducting high quality writing intervention research. Each recommendation includes multiple suggestions and guidelines for designing, conducting, and presenting a strong study. The confidence that can be placed in the findings of a writing intervention study depends on how well a researcher addresses these recommendations, suggestions, and guidelines. A checklist in the Appendix brings all of these together, providing a checklist for designing, conducting, and reporting a high quality writing intervention study. We did not include suggestions for the 12th recommendation on designing a series of studies in this Table, as it is intended to serve as a guide for conducting an individual investigation.

In closing, we make several additional observations. First, conducting a study or series of studies to determine if a writing practice is effective is only a first step. One of the ultimate goals of writing intervention research is to improve the teaching of writing (another goal is to inform theory). There is no guarantee, however, that a writing practice that was effective in a series of research studies will be effective in all other situations. Frankly, there is never a perfect match between the conditions under which a writing practice was tested by researchers and the conditions in which it is subsequently applied by those who teach writing. Consequently, additional research is needed to determine how to transfer evidence-based practices into everyday school practices. Otherwise, potentially promising practice for teaching writing will remain on the shelf so to speak, admired but not used.

Second, research conducted with teachers successfully delivering the writing intervention are especially important, as such investigations demonstrate that a practice can be delivered under relatively normal conditions. Again though, if a writing practice is to be applied more broadly, we need to understand better the context in which this is to occur. This is no easy task, as there are many levers as well as constraints that determine if a practice is used initially and over time by teachers. This includes the ways that institutional factors at the state, community, school, and classroom-level shape the teaching of writing as well as the ways that teachers and students' experiences, expectations, values, and motive determine what happens in the classroom. For instance, if a teacher does not value a specific writing practice, views it

as too complex, or judges that it is too time consuming, it is unlikely to be used regardless of how effective it was in one or more research studies.

Finally, many of the recommendation and points made in the article and summarized in the Appendix are pertinent to intervention research in domains other than writing. Most of them also apply to other types of quantitative research as well.

Seasoned researchers are likely to view our recommendations, suggestions, and guidelines presented here as "givens." We hope so. However, in examining writing intervention literature (e.g., Graham et al., 2012; Graham & Perin, 2007; Rogers & Graham, 2008), we were struck by the fact that it is not as good as it can or should be. While there are many very good writing intervention studies, too many are of poor quality.

References

Achieve, Inc. (2005). *Rising to the challenge: Are high school graduates prepared for college and work?* Washington, DC: Author.

Applebee, A., & Langer, J. (2011). A snapshot of writing instruction in middle and high schools. *English Journal, 100,* 14-27.

Berninger, V., Fuller, F., & Whitaker, D. (1996). A process model of writing development across the life span. *Educational Psychology Review, 8*, 193-218. http://dx.doi.org/10.1007/BF01464073

Berninger, V., Mizokawa, D., & Bragg, R. (1991). Theory-based diagnosis and remediation of writing disabilities. *Journal of School Psychology, 29*, 57-79.http://dx.doi.org/10.1016/ 0022-4405(91)90016-K.

Boldrini, E. & Cattaneo, A. (2012). Collaborative writing on professional experiences as a mean for fostering reflective thinking in the VET curricula. In C. Gelati, B. Arfe, & L. Mason (Eds.), *Issues in writing research: In honor of Piero Boscolo.* Padova, Italy: University of Padova.

Braaksma, M.A.H., Rijlaarsdam, G., & Van den Bergh, H. (2014). Differential effects of learning conditions on content knowledge acquisition and writing skills. Two experimental studies on the effects of written production and learning modes. *Learning & Instruction.*

Brindle, M. (2012). *Examining relationships among teachers' preparation, efficacy, and writing practices* (Unpublished doctoral dissertation). Vanderbilt University, Nashville, TN.

Bryk, S., & Raudenbush, A.. (2002). *Hierarchical linear models : applications and data analysis methods* (2. ed.). Thousand Oaks, CA: Sage Publications

Campbell, D. & Stanley, J. (1963). Experimental and quasi-experimental designs for research. Chicago, IL: Rand-McNally.

Camps, A., & Milan, M. (2000). *Metalinguistic activity in learning to write.* Amsterdam: Amsterdam University Press.

Common Core State Standards: National Governors Association and Council of Chief School Officers (2010). Downloaded from: http://*www.corestandards.org/*.

Cook, B., Smith, G., & Tankersley, M. (2012). Evidenced-based practices in education. In. K.R. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook* (Vol. 1, pp. 495-528). Washington, D.C.: APA.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich.

Festas, I, Oliveira, A., Rebelo, J, Damião, M., Harris, K.R., & Graham, S. (in press). The Effects of Self-Regulated Strategy Development (SRSD) on the Writing Performance of Eighth Grade Portuguese Students. *Contemporary Educational Psychology*.

Fitzgerald, J., & Markham, L. (1987). Teaching children about revision in writing. *Cognition and Instruction, 4*, 3-24. http://dx.doi.org/10.1207/s1532690xci0401_1.

Fitzgerald, J., & Teasley, A. (1986). Effects of instruction in narrative structure on children's writing. *Journal of Educational Psychology, 78*, 424-432.http://dx.doi.org/10.1037/0022-0663.78.6.424.

Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in grades 4 to 6: A national survey. *Elementary School Journal, 110,* 494-518. http://dx.doi.org/10.1086/651193.

Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology* (pp. 457-478). Mahwah, NJ: Erlbaum.

Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities: A review. *Learning Disability Quarterly, 22*, 78-98. http://dx.doi.org/10.2307/1511268.

Graham, S. (1990). The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology, 82,* 781-791.http://dx.doi.org/10.1037/0022-0663.82.4.781

Graham, S., & Harris, K.R. (in press). Evidence-based practice and writing instruction. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (Volume 2). NY; Guilford.

Graham, S., Harris, K.R. (2012). The role of strategies, knowledge, will, and skills in a 30 year program of research (with homage to Hayes, Fayol, and Boscolo). In V. Berninger (Ed.), *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology* (pp. 177-196). London: Psychology Press.

Graham, S., & Harris, K.R. (2009). Almost 30 Years of Writing Research: Making Sense of It All with *The Wrath of Khan. Learning Disabilities Research & Practice, 24,* 58-68. http://dx.doi.org/10.1111/j.1540-5826.2009.01277.x

Graham, S., & Harris, K.R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the Center to Accelerate Student Learning. *Journal Special Education, 39*, 19-33. http://dx.doi.org/10.1177/00224669050390010301

Graham, S., & Harris, K.R. (1997). It can be taught, but it does not develop naturally: Myths and realities in writing instruction. *School Psychology Review, 26*, 414-424.

Graham, S., & Harris, K. R. (1989). A components analysis of cognitive strategy instruction: Effects on learning disabled students' compositions and self-efficacy. *Journal of Educational Psychology, 81*, 353-361.http://dx.doi.org/10.1037/0022-0663.81.3.353

Graham, S., Harris, K.R., & Fink, B. (2000). Is handwriting causally related to learning to write? Treatment of handwriting problems in beginning writers. *Journal of Educational Psychology,92*, 620-633. http://dx.doi.org/10.1037/0022-0663.92.4.620

Graham, S., Harris, K. R., & Fink-Chorzempa, B. (2002). Contributions of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology, 94*, 669-686.http://dx.doi.org/10.1037/0022-0663.94.4.669

Graham, S., Harris, K.R., & Hebert, M. (2011a). *Informing Writing: The benefits of formative assessment*. Alliance for Excellence in Education. Washington, D.C.

Graham, S., Harris, K.R., & Hebert, M. (2011b). It is more than just the message: Analysis of presentation effects in scoring writing. *Focus on Exceptional Children, 44* (4), 1-12.

Graham, S., Harris, K.R., & Mason, L. (2005). Improving the writing performance, knowledge, and motivation of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology, 30*, 207-241.http://dx.doi.org/10.1016/j.cedpsych.2004.08.001

Graham, S., Harris, K. R., & McKeown, D. (2013). The writing of students with LD and a meta-analysis of SRSD writing intervention studies: Redux. In L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of Learning Disabilities* (2nd Ed., pp. 405-438). New York, NY: Guilford Press.

Graham, S., Harris, K.R., & Santangelo, T. (in press). Research-based writing practices and the Common Core: Meta-analysis and meta-synthesis. *Elementary School Journal*.

Graham, S., Harris, K.R., & Zito, J. (2005). Promoting internal and external validity: A synergism of laboratory experiments and classroom-based research. In G. Phye, D. Robinson, & J. Levin

(Eds.), *Empirical methods for evaluating educational interventions* (pp. 235-265). San Diego, CA: Elvieser.

Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81,* 710-744.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104,* 879-896.http://dx.doi.org/10.1037/a0029185

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99,* 445-476.http://dx.doi.org/10.1037/0022-0663.99.3.445

Harris, K.R. (1982). Cognitive-behavior modification: Application with exceptional students. *Focus on Exceptional Children, 15*(2), 1-16.

Harris, K.R. (1985). Conceptual, methodological, and clinical issues in cognitive-behavioral assessment. *Journal of Abnormal Child Psychology, 13*, 373-390.http://dx.doi.org/10.1007/BF00912723

Harris, K., & Graham, S. (1985). Improving learning disabled students' composition skills: Self-control strategy training. *Learning Disability Quarterly, 8*, 27-36. http://dx.doi.org/10.2307/1510905

Harris, K.R., & Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learning Disability Quarterly, 22*, 251-262.http://dx.doi.org/10.2307/1511259

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

Harris, K.R., & Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learning Disability Quarterly, 22*, 251-262. http://dx.doi.org/10.2307/1511259

Harris, K.R., & Graham, S. (2013). Integrating reading and writing instruction In B. Miller, P. McCardle, & R. Long (Eds.), *Teaching reading and writing: Improving instruction and student achievement*. Baltimore, MD: Paul H. Brookes Publishing Co.

Harris, K.R., Graham, S., & Adkins, M. (in press). Practice-based professional development and self-regulated strategy development for Tier 2, at-risk writers in second grade. *Contemporary Educational Psychology*.

Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and children's writing. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 131-153). Mahwah, NJ: Erlbaum.

Harris, K.R., Graham, S., & Mason, L. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43,* 295-340.http://dx.doi.org/10.3102/00028312043002295

Harris, K. R., Graham, S., Mason, L., & Friedlander, B. (2008). *Powerful writing strategies for all students*. Baltimore, MD: Brookes Publishing.

Harris, K.R., Lane, K.L., Driscoll, S., Graham, S., Wilson, K., Sandmel, K., Brindle, M., & Schatschneider, C. (2012a). Tier 1, teacher-implemented self-regulated strategy development for students with and without behavioral challenges. *Elementary School Journal, 113,* 160-191.http://dx.doi.org/10.1086/667403

Harris, K.R., Lane, K., Graham, S., Driscoll, S., Sandmel, K., Brindle, M., & Schatschneider, C. (2012b) Practice-based professional development for strategies instruction in writing: A randomized controlled study. *Journal of Teacher Education, 63,* 103-119.http://dx.doi.org/10.1177/0022487111429005

Harris, K.R., & Pressley, M. (Guest Editors). (1994). Integrated approaches to instruction and knowledge construction. Special issue of *Reading and Writing Quarterly: Overcoming Learning Difficulties, 10*.

Hayes, J. (2011). Kinds of knowledge-telling: Modeling early writing development. *Journal of Writing Research, 3*, 73-92.

Hayes, J. (2012). Modeling and remodeling writing. *Written Communication, 29,* 369-386. http://dx.doi.org/10.1177/0741088312451260

Hayes, J., & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3 -30). Hillsdale, NJ: Erlbaum.

Hebert, M., Graham, S. Rigby-Wills, H., Ganson, K. (in press). Effects of Writing on Informational Text Comprehension: An Experimental Study with Fourth Grade Students. *Learning Disabilities: A Contemporary Journal*.

Hillocks, G. (2006). The testing trap: How state writing assessments control learning. New York, NY: Teachers College Press.

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. http://dx.doi.org/10.1177/001440290507100203

Kiuhara, S., Graham, S., & Hawken, L. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology, 101,* 136-160.http://dx.doi.org/10.1037/a0013097

Harris, K.R., Lane, K., Driscoll, S., Graham, S., Wilson, K., Sandmel, K., Brindle, M., Schatschneider, C. (2012). Tier 1. teacher-implemented self-regulated strategy development for students with and without behavior concerns. *Elementary School Journal, 113,* 160-191.http://dx.doi.org/10.1086/667403

Lane, K., Harris, K.R., Graham, S., Driscoll, S., Sandmel, K., Morphy, P., Hebert, M., & House, E. (2011). The effects of self-regulated strategy development for second-grade students with writing and behavioral difficulties: A randomized controlled trial. *Journal of Research on Educational Effectiveness*, *4,* 322-353.http://dx.doi.org/10.1080/19345747.2011.558987

Larzelere, R., Kuhn, B., & Johnson, B. (2004). The intervention selection bias: An underrecognized confound in intervention research. *Psychological Bulletin, 130,* 289-303. http://dx.doi.org/10.1037/0033-2909.130.2.289

Leijten, M., Van Waes L., Schriver, K., & Hayes, J.R. (2014). Writing in the workplace: constructing documents using multiple digital sources. Journal of Writing Research, 5, 285-337.

Levin, J. (1985). Some methodological and statistical "bugs" in research on children's learning. In M. Pressley & C. Brainerd (Eds.), *Cognitive learning in children* (pp. 205-233). NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4613-9544-7_7

Levin, J. (2005). Randomized control trials on trial. In G. Phye, D. Robinson, & J. Levin (Eds.), *Empirical methods for evaluating educational interventions* (pp. 3-27). San Diego, CA: Elvieser.

Lipsey, M., & Wilson, D. (1993). The efficacy of psychological. Educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181-1209. http://dx.doi.org/10.1037/0003-066X.48.12.1181

MacArthur, C., & Graham, S. (1987). Learning disabled students' composing under three methods of text production: Handwriting, word processing, and dictation. *Journal of Special Education, 21*, 22-42.http://dx.doi.org/10.1177/002246698702100304

MacArthur, C., Schwartz, S., Graham, S., Molloy, D., & Harris, K.R. (1996). Integration of strategy instruction into a whole language classroom: A case study. *Learning Disabilities Research and Practice, 11,* 168-176.

McCutchen, D. (1988). "Functional automaticity" in children's writing: A problem of metacognitive control. *Written Communication, 5*, 306-324. http://dx.doi.org/10.1177/0741088388005003003

McKeown, D. (2012). Examining the effects of practice-based professional development and coaching on the differentiation of SRSD writing instruction (Unpublished doctoral dissertation). Vanderbilt University, Tennessee.

National Center for Education Statistics (2012).*The Nation's Report Card: Writing 2011* (NCES 2012–470). Institute of Education Sciences, U.S. Department of Education, Washington,

National Council of Teachers of English. (2004). NCTE beliefs about the teaching of writing. Retrieved from http://www.ncte.org/print.asp?id= 118876&node=367

Pressley, M., Graham, S., & Harris, K.R. (2006). The state of educational intervention research. *British Journal of Educational Psychology, 76,* 1-19. http://dx.doi.org/10.1348/000709905X66035

Pressley, M. & Harris, K.R. (Guest Editors). (1994a). Increasing the quality of educational intervention research. Special issue of *Educational Psychology Review, 6*.

Pressley, M., & Harris, K.R. (1994b). More about increasing the quality of educational intervention research: A synthesis. *Educational Psychology Review, 6*, 191-208.http://dx.doi.org/10.1007/BF02213181

Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E., & Reedts, M. (2012). Writing. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook* (Volume 3; pp. 189 – 228). Washington, D.C.: American Psychological Association.

Rogers, L., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology, 100,* 879 - 906.http://dx.doi.org/10.1037/0022-0663.100.4.879

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology, 97*, 43-54.http://dx.doi.org/10.1037/0022-0663.97.1.43

Sandmel, K., & Graham, S. (2011). The process writing approach: A meta-analysis. *Journal of Educational Research, 104,* 396-407.http://dx.doi.org/10.1080/00220671.2010.488703

Sawyer, R., Graham, S., & Harris, K. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on learning disabled students= composition skills and self-efficacy. *Journal of Educational Psychology, 84*, 340-352.http://dx.doi.org/10.1037/0022-0663.84.3.340

Scardamalia, M., & Bereiter, C. (1986). Written composition. In M. Wittrock (Ed.). *Handbook of research on teaching* (3rd Ed., pp. 778-803). New York: MacMillan.

Sheils, M. (1975, December 8). Why Johnny can't write. *Newsweek,* p. 58.

Smagorinki, P. (1987). Graves revisited: A look at the methods and conclusions of the New Hampshire study. *Written Communication, 4,* 331-342.http://dx.doi.org/10.1177/0741088387004004001

Stein, N., & Glenn, C. (1978). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Advances in discourse processes: Volume 2. New Directions in discourse processing.* Norwood, NJ: Ablex.

Stinner, D., & Tennent, D. (2012). Losses to follow-up present a risk to study validity: Differential attrition can be a shortcoming in clinical research. *AAOS Now* (February). http://www.aaos.org/news/aaosnow/feb12/research1.asp

Swanson, L., Harris, K.R., & Graham, S. (2013). *Handbook of Learning Disabilities* (2nd Ed., pp. 405-438). New York, NY: Guilford Press.

Van Steendam, E., Rijlaarsdam, G., Van den Bergh, H., Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *Instructional Science*, art.nr. May 2014.

Van Steendam, E., Tillema, M., Rijlaarsdam, G., & van den Bergh, H. (2012). Measuring writing: Recent insights into theory, methodology and practice (Studies in Writing Series). Leiden: Brill Books.

Appendix

Checklist for planning, conducting, and reporting a high quality writing intervention research study

The study

- □ addresses an important theoretical, empirical, and/or
- □ practical problem
- □ includes clearly stated research questions
- □ includes hypothesis for each research question.
- □ is designed and carried out in an ethical manner

The writing intervention

- □ is theoretically situated
- □ draws on previous findings in the writing research literature
- □ based on an analyses of the writing task, learners, context, and desired outcomes
- □ applies effective instructional methods
- □ is piloted (and refined based on the pilot)
- □ is described in enough detail to be replicable.

Instructors/teachers administering the writing intervention

- □ are provided with lessons for implementing it
- □ are taught how to apply it
- □ administer it with fidelity during the study.

The comparison/control condition

- □ is credible
- □ is similar to the intervention except for features that distinguish it by definition
- □ developed with the same level and care as the writing intervention (see above)
- □ is piloted (and refined based on the pilot)
- □ described in enough detail to be replicable
- □ is administered with fidelity (data is collected to describe what occurred).

Each assessment measure

- □ is reliable
- □ is valid for the purpose to which they are applied
- □ does not exhibit floor/ceiling problems
- □ is appropriate for the participants in the study
- □ include equivalent forms (if testing is to occur at more than one time point)
- □ includes enough samples to provide a valid estimate of performance.

Assessment procedures evaluate
- ☐ what was taught
- ☐ more generalized improvement
- ☐ maintenance over time.

Test administrators
- ☐ are blind to the purpose/design of the study
- ☐ follow a protocol for administering each measure
- ☐ are trained to apply these protocols
- ☐ reliably administer these protocols during the study
- ☐ deliver measures in a counterbalance order at and across time points.

The resulting data for each measure is scored after
- ☐ all identifying information is removed
- ☐ papers are typed and pertinent errors corrected (if the measure involves writing)
- ☐ they are randomly ordered.

The context within which the study occurs
- ☐ are randomly sampled
- ☐ is similar to normal conditions (when random sampling is not possible)
- ☐ are adequately described.

Participants in the study
- ☐ are randomly sampled
- ☐ are adequately described (this includes both students and instructors/teachers).

The design of the study
- ☐ provides a suitable means for answering the proposed research questions
- ☐ includes random assignment of students to conditions (i.e., true-experiment)
- ☐ involves matching students in each condition on relevant variables (for quasi-experiments)
- ☐ includes pretests (except for true-experiments of short duration).

The design of the study controls for confounds involving
- ☐ selection bias (i.e., pre-instructional differences between participants in each condition)
- ☐ teacher effects
- ☐ instructional time between conditions
- ☐ types of writing completed by participants in each condition
- ☐ expectations of participants in each condition (e.g., Hawthorne and John Henry effects
- ☐ drift in the application of the writing treatment or control/comparison conditions

- contamination (features of the writing intervention applied in control/comparison students
- assigning a single teacher to each condition (i.e., one teacher, one condition)
- attrition.

The statistical procedures

- are suitable for the design of the study and for answering the proposed research questions
- are determined in advance (and maintained unless there are good reasons for deviation)
- include a power analysis to determine sample size
- meet the assumptions on which they are based
- involve the proper unit of analysis
- include appropriate adjustments in significance levels (e.g., a Bonferroni adjustment).

The report describing the study

- theoretically situates the study and the resulting findings
- makes it clear how the study and resulting findings replicate and extend previous research
- describes how study procedures were implemented and sequenced
- presents pertinent statistics for each measure (M, SD, CI, correlations, effect size)
- draws conclusions that are consistent with the data and do not go beyond it
- discusses possible reasons for unexpected findings
- suggests directions for future research
- presents implications for practice.

Note: *M* = mean; *SD* = standard deviation; *CI* = confidence interval