2014-12-01

Evaluating the impact of feedback on elementary aged students' fluency growth in written expression: a randomized controlled trial.

Adrea J Truckenmiller, Tanya L Eckert, Robin S Codding and Yaacov Petscher

Published in final edited form as:
J Sch Psychol. 2014 December ; 52(6): 531–548. doi:10.1016/j.jsp.2014.09.001.

Evaluating the Impact of Feedback on Elementary Aged Students' Fluency Growth in Written Expression: A Randomized Controlled Trial

Adrea J. Truckenmiller, **Tanya L. Eckert**, **Robin S. Codding**, and **Yaacov Petscher**
Syracuse University, University of Massachusetts-Boston, Florida State University

Adrea J. Truckenmiller: atruckenmiller@fcrr.org; Tanya L. Eckert: taeckert@syr.edu; Robin S. Codding: robin.codding@umb.edu; Yaacov Petscher: ypetscher@fcrr.org

Abstract

The purpose of this randomized controlled trial was to evaluate elementary-aged students' writing fluency growth in response to (a) instructional practices, (b) sex differences, and (c) student's initial level of writing fluency. Third-grade students (n=133) in three urban elementary schools were randomly assigned to either an individualized performance feedback condition (n=46), a practice-only condition (i.e., weekly writing practice; n = 39), or an instructional control condition (n = 48) for 8 weeks. Findings included support for use of performance feedback as an instructional component in general education classrooms (Hedges' g = 0.66), whereas simple practice with curriculum-based measurement in written expression did not produce growth significantly greater than standard instructional practices. The hypothesis that girls write significantly more than boys was supported. However, girls and boys did not differ in their rate of growth. Finally, students' initial risk status in writing fluency did not differentially predict growth in writing fluency over the course of the study. Implications for incorporating feedback as a basic component of intervention in writing are discussed.

Many schools use data from specific assessment tools, such as curriculum-based measurement (CBM; Deno, 2003), to make instructional decisions regarding basic academic skills in reading, math, and writing. From a pragmatic standpoint, educational decisions regarding the effectiveness of school-based interventions are generally made over short time periods. For example, response to intervention practices recommends evaluating progress monitoring data to gauge the success of interventions within 6 to 10 weeks of implementation (Fuchs & Fuch, 2006; Gresham, 2007). Given this decision-making process and the lack of information on basic skills in writing, relative to reading and math, it is important to explore and understand (a) the impact of interventions on growth in basic skills of written expression, (b) the expected rate of growth in written expression over short intervals, and (c) the variables that affect growth in written expression. The current study aims to address each of these areas in order to increase the research base on instructional practices for writing and inform data-based decision-making by school professionals. These aims are consistent with the national demand for evidence-based interventions and the

Corresponding author at: Florida Center for Reading Research 2010 Levy Ave, Suite 100, Tallahassee, FL 32310, USA. Tel.: +1 850 645 2474.

practical considerations that arise in school settings (American Psychological Association Presedential Task Force on Evidence-Based Practice, 2006; Kratochwill & Shernoff, 2003/2004; Stoiber & Kratochwill, 2000). Furthermore, this type of information regarding written expression is of practical importance for school psychologists as concerns about written expression are second only to reading problems in number of referrals that school psychologists receive nationally (Bramlett, Murphy, Johnson, & Wallingsford, 2002).

Writing instruction and intervention

Although there is a general consensus about the broad topics covered in general education writing instruction, curricula across states, school districts, and even classrooms within the same school are highly varied (Berninger et al., 2006; Cutler & Graham, 2008; Graham, Harris, MacArthur, & Fink, 2002). For example, 65% of a sample of primary grade teachers indicated that they did not use any commercial programs to teach writing, and the remaining 35% reported using 137 different programs (Cutler & Graham, 2008). This high level of variation in writing instruction is not surprising given the complexity of the skill and the relative lack of research and evidence to support instructional practices in writing compared with reading (Graham, Harris, & Hebert, 2011).

The effects of current instructional practices in the United States can be most directly observed by examining the recent reports of the National Assessment of Educational Progress indicating that a substantial percentage of school-aged students have not demonstrated mastery of writing skills considered fundamental for proficient work. Specifically, 72% of fourth-grade students (Persky, Daane, & Jin, 2003), 74% of eighth-grade students, and 73% of twelfth-grade students could not write at the Proficient Level for their grade level (National Center for Education Statistics, 2012). These findings further substantiate national policy reports on the condition of writing in U.S. public schools (National Commission on Writing, 2003), wherein writing has been characterized as the “the neglected R.”

Writing development

One influential theoretical model of writing development was proposed by Hayes and Flower (1980) and emphasizes the cognitive factors that impact the writing process and overall development of writing skills. In this model, Hayes and Flower specify that writing is a recursive process that comprises three components: (1) planning (i.e., idea generation, organization, goal setting), (2) translating (i.e., transforming representations of language into text), and (3) reviewing the written product (i.e., evaluating and revising the written product). Berninger et al. (1997) argued that although the model developed by Hayes and Flower may be appropriate for conceptualizing written expression in adults, it does not take into account developmental processes, in particular the significant role that transcription plays in children developing emergent writing skills. Transcription refers to the combination of syntactic, lexical, phonological, and motor tasks that need to be automatized in short term memory and identified as translation by Hayes and Flower. Berninger and colleagues argued that because beginning writers have not yet mastered the lower-level process of producing text, they are unable to attempt the more complex processes of planning and reviewing that

Hayes and Flower proposed are more critical to continual improvement of written products. A theoretical model was thus proposed that underlies the importance of transcription in beginning writers (Abbott & Berninger, 1993; Berninger et al., 2002). This model, the Simple View of Writing, demonstrates that transcription is necessary (but not sufficient) for the broader goal of composition.

Abbott and Berninger (1993) demonstrated that the developmental progression in writing has two primary steps: transcription (i.e., retrieval from memory of orthographic symbols that could then be coordinated with motor output of those symbols) followed by text generation (i.e., generating ideas and placing the words into grammatically-correct expressions). Transcription skills need to develop first and generally constitute the focus of writing instruction from kindergarten through second grade, in which spelling, handwriting and orthographic skills (i.e., transcribing sounds and language into symbols that compose written language) are targeted. The second step of the developmental progression, text generation, generally begins in third grade and continues throughout formal schooling. This step involves the development and accommodation of writing skills in extended text composition for increasingly broader applications. Ultimately, mastery of transcription in the early grade levels is essential for future writing success in successive grade levels (Berninger et al., 2006).

Both Hayes and Flower's (1980) model and the Simple View of Writing (Berninger et al., 2002) are used to inform instructional efforts among writers of varying skill levels. However, because the Simple View of Writing proposed by Berninger and colleagues (2002) accounts for the development of emergent writing skills, their model is used to inform instructional efforts among elementary-aged children. For example, a number of handwriting and spelling intervention studies were developed for emerging writers in the primary grades based on the Simple View of Writing's assumption that transcription emerges first (e.g., Berninger et al., 1997; Jones & Christenson, 1990). As children begin developing transcription skills in the third and fourth grades, additional practice composing extended text should be provided (Berninger et al., 2006) in order to develop automatic and flexible transcription processes in their written compositional skills (Eckert, Codding, Truckenmiller, & Rheinheimer, 2009). The model proposed by Hayes and Flower, in contrast, tends to inform interventions (such as self-regulated strategy instruction) that focus on improving students' writing outcomes who have developed transcription processes yet need additional assistance in planning and reviewing their compositions. According to three recent meta-analyses (Graham & Perin, 2007; Graham, Harris, & Fink-Chorzempa, 2003; Graham, Harris, & Hebert, 2011), self-regulated strategy in writing composition was found to be effective in improving the writing quality and fluency of middle- and high-school students who acquired basic transcriptional skills.

Focusing on transcription skills, a number of studies (Deno, 2003; Deno, Mirkin, & Marston, 1980; Van Houten, Morrison, Jarvis, & McDonald, 1974) documented that increases in elementary-aged students' rate of writing (i.e., writing fluency) are associated with improvements in general writing quality, including performance on high-stakes tests (Powell-Smith & Shinn, 2004). The target outcome of the present study is writing fluency, which is defined as a quantitative measure indicating the amount of text produced within time constraints (Deno, Marston, & Mirkin, 1982). This measurement of writing fluency, referred to as curriculum-based measurement in written expression (CBM-WE), results in an index of student performance that has the added benefit of being sensitive to changes over time (Marston, Deno, & Tindal, 1983).

Interventions targeting writing fluency

Interventions targeting fluency (e.g., goal-setting, performance feedback, and tangible reinforcement) in basic academic skills have been found to be effective as supplements to a wide variety of core instructional practices (Eckert et al., 2009; Griffiths, Van Der Heyden, Parsons, & Burns, 2006). Fluency-building interventions are conceptualized as time-efficient practices that can supplement instruction or intervention already occurring in the classroom (Martens, Daly, Begeny, & Van Der Heyden, 2011). Empirical support for these interventions suggests moderate to strong effects (Eckert et al., 2009; Morgan & Sideridis, 2006), although most fluency-building interventions have focused on reading and mathematics (e.g., Codding et al., 2007; Morgan & Sideridis, 2006). The existing literature base demonstrating the efficacy of fluency-based writing interventions in general education classrooms is small and focuses predominately on one type of fluency-building intervention: performance feedback (Eckert et al., 2009).

A performance feedback intervention with writing was originally examined in three studies conducted by Van Houten and colleagues (Van Houten, 1979; Van Houten et al., 1974; Van Houten, Hill, & Parsons, 1975) with positive results for writing fluency. Clinical utility of these procedures was also demonstrated by documenting that teachers could easily implement the interventions on their own (Van Houten, 1979). Although this system was found to be effective, it is unclear whether performance feedback, in isolation, was responsible for the changes in the students' behavior, as any number of components (e.g., self-scoring and public posting) interacting in the system utilized could be responsible for the positive effects. Among others, Kazdin (2008) has pointed out that very few intervention studies specifically focus on isolating the mechanisms of behavior change, despite the importance of this information for improving intervention effectiveness.

Recently, the effectiveness of using performance feedback, in isolation, as an intervention for writing fluency was explored (Eckert et al., 2006). The intervention is simple to implement and can be used as a fluency-building intervention to supplement core instructional programs used in the classroom. It can be readily administered with an entire classroom of students (i.e., as a universal intervention) or with groups or individual students. Results from two studies of elementary-aged children (Eckert et al., 2006; Eckert, Truckenmiller, Rheinheimer, Perry, & Koehler, 2008) indicated that individualized performance feedback on a brief CBM-WE probe delivered once per week has moderate to high effects ($d = 0.76$; Eckert et al., 2006) on children's writing fluency growth (as measured by total words written) and additionally, that practice with CBM-WE may also have some positive effect on writing fluency growth.

Although effectiveness and utility of this fluency-based intervention has been demonstrated, a randomized control trial has not been conducted. To date, no study has compared rates of writing fluency change following performance feedback to a condition where students are provided with no regular writing practice beyond typical classroom instruction. Similarly, the effect of writing practice alone (i.e., without intervention) on the trajectory of elementary-aged student's writing fluency has not been systematically evaluated. As a result, it is unclear whether students' writing growth may benefit simply from practice with CBM-WE probes. Experts have suggested that students may just need the opportunity to get enough repetition (i.e., practice) in writing to become fluent (Abbott & Berninger, 1993; Graham & Harris, 1997). Providing time for students to practice writing on a daily basis is one of several evidence-based recommendations published in a What Works Clearinghouse practice guide for teaching elementary schools students to be effective writers (Graham et al., 2012).

Given the wide variety of instructional practices used for writing in elementary schools (Cutler & Graham, 2008), growth in writing fluency in the absence of supplemental intervention may also be highly variable. Several studies have demonstrated significant positive growth across the elementary school years, with one longitudinal study showing growth of 1.65 standard deviations (Hedges' g = 1.65) from first grade to third grade (Coker, 2006). Although growth in writing fluency is significant, the amount of growth is relatively small across a large period of time. Only one longitudinal study has demonstrated preliminary evidence that growth occurs between the fall and spring of an academic year (partial $\eta^2 = 0.13$ to 0.15; Malecki & Jewell, 2003). When frequently measured across a time period shorter than a school year, it may be difficult to detect that small amount of growth. For example, Marston, Lowry, Deno, and Mirkin (1981) reported that less than half of elementary-aged students evidenced increased writing fluency performance from fall to spring. One set of national norms exists for describing weekly growth in CBM-WE. Specifically, AIMS-web (Pearson Education, Inc., 2009) estimates indicate that third grade students' rate of improvement from fall to winter to spring is 0.4 units (i.e., total words written, words spelled correctly, and correct writing sequences) per instructional week.

Variables affecting writing fluency

In addition to instructional and intervention variables, there are less conspicuous factors that may influence students' growth in writing fluency, including sex, socioeconomic status, initial level of fluency, special education designation, and teacher variables (Coker, 2006). Two factors, sex differences and initial level of fluency have been investigated in a number of studies of educationally important outcomes are explored further in this study.

Sex differences

Overall, boys have more difficulties in general writing performance. National achievement data indicates that elementary-aged girls (Persky et al., 2003), intermediate-aged girls, and high school-aged girls (Salahu-Din et al., 2008) all significantly outperform boys on general writing proficiency as measured by the National Assessment of Educational Progress. Boys are also at higher risk of being diagnosed with a learning disability in writing (Berninger, Nielson, Abbott, Wijsman, & Raskind, 2008) and perform significantly lower than girls on frequently-used norm-referenced assessment subtests, including spelling, capitalization, punctuation, and language usage (Martin & Hoover, 1987).

In general academic achievement, no sex differences are evident on untimed achievement tests in reading and mathematics (Camarata & Woodcock, 2006). However, on the corresponding timed fluency tests, female students outperform male students indicating that females have higher performance on timed tests only. Interestingly, in the area of writing, female students significantly outperformed male students on an untimed test of writing achievement (d = 0.33) and a timed test of writing fluency (d = 0.42; Camarata & Woodcock, 2006). Others have examined writing fluency as it relates to sex and demonstrated that the gap between girls' and boys' performance is even greater when students are given a longer amount of time to write (i.e., 5 min versus 3 min; Berninger et al., 2008).

Given the differences between boys' and girls' writing fluency, Berninger and Fuller (1992) suggested that outcome analyses be separated by sex. In their large sample of third grade students, they reported that girls wrote an average of 51.5 words and boys wrote an average 45.7 words in a 5-minute span. A few studies have examined sex differences in writing fluency as measured by several metrics in CBM-WE (Fearrington et al., 2014; Kim, Al Otaiba, Wanzek, & Gatlin, 2014; Malecki & Jewell, 2003). Kim et al. (2014) found significantly lower performance by boys than girls on a variety of writing outcomes (i.e., writing quality, production, and CBM-WE), with a 0.37 standard deviation difference in performance on CBM-WE measures favoring girls. Within their sample of 481 third-, fourth-, and fifth-grade students, Malecki and Jewell (2003) reported that girls outperformed boys, writing 44.7 total words written (TWW), 42.1 words spelled correctly (WSC), and 40.2 correct writing sequences (CWS) in a period of 3 min compared with 38.6 TWW, 35.5 WSC, and 33.1 CWS. The authors reported these differences to represent a small to medium effect size. Sex differences accounted for about 5% of the total variance in TWW, WSC, and CWS. An interaction between sex and age (i.e., grade level) was also found. That is, the gap between girls and boys widened across the three developmental periods used by the researchers (early elementary, elementary, and middle school) when examining performance on production-dependent measures (i.e., TWW, WSC, and CWS). Alternately, the gap decreased with increasing age on production-independent measures (i.e., percentage of correctly spelled words and percentage of correct writing sequences). However, this interaction was not explicitly tested at the longitudinal level (i.e., using a within-subject design). Finally, Fearrington et al. (2014) examined the differences in boys and girls performance in TWW, WSC, and CWS at three time points in grades 3 through 8. They also found significant differences between boys and girls at each time point and all grade levels. They found a significant interaction and medium effect size (ES = 0.08) between gender and time for grade 3 only. Again the improvements in CBM-WE over time favored girls. Their results showed that some sex differences were partially explained by language and cognitive skill differences, but not all of the writing outcome differences could be explained by the variables included in their study.

Although it is likely that the sex differences in students' writing growth over time generally occur, this conclusion may be limited by study methodology (i.e., cross-sectional design or limitations with ANOVA; Raudenbush, 2001). Given the limited empirical data available describing the effect of sex on elementary-aged students' writing fluency growth from a longitudinal perspective, it is reasonable to predict that there are significant sex differences in elementary-aged students' level of writing fluency. It is unclear however, whether sex differences in elementary-aged students' writing fluency growth (i.e., slope) exist.

Initial level of writing fluency

Regardless of the factors (e.g., sex) that explain variability in initial level of writing fluency, it can be argued that it is just as important to identify how a student's initial level of fluency predicts growth over time. For example, a school-based practitioner will likely expect a different rate of growth in writing fluency from a student writing at an instructional level than from a student whose performance is at the frustrational level. In an early study of CBM-WE, developers examined the ways in which CBM-WE performance differed between low-achieving students and students with a diagnosed learning disability (Shinn, Ysseldyke, Deno, & Tindal, 1982). The two groups did not differ significantly on a standardized measure of writing; however, the students with learning disabilities did produce lower writing fluency rates (M = 37.6 TWW) than the low-achieving students (M = 43.9 TWW). The students who were classified with a learning disability gained an average of 1.04 words per week over a 5-week period, whereas the low-achieving students actually demonstrated decreases in their writing fluency growth (i.e., 0.53 words per week), representing a statistically significant difference in growth. It is possible that the reported findings were due to curricular differences (i.e., general education versus special education) or measurement issues associated with initial fluency levels (i.e., floor and ceiling effects as well as lack of statistical modeling of measurement error). These results suggest that growth expectations may differ based on a student's initial fluency level.

Although initial status has not been explored heavily in the area of writing, response to intervention practices utilizes a norm-based or benchmark-based cut point on screening measures to determine which students may benefit from additional small group instruction or intervention in a particular skill (Fuchs, 2003). There are a variety of ways that are recommended for choosing cut points; however, very few studies have explicitly studied initial fluency level in the area of CBM-WE. The most recent set of national norms is provided by AIMSweb, which reports metrics at the 25th percentile for each grade level.

Purpose of the current study

The primary aim of this study was to examine the impact of basic instructional practices (i.e., practice, performance feedback, and typical instructional practices) on commonly used CBM-WE outcome metrics (i.e., TWW and CWS) during a brief intervention period (i.e., progress monitoring period). Growth in writing fluency due to typical instructional practices has been shown to be significant but small across a school year (Coker, 2006; Malecki & Jewell, 2003). However, it is unknown if changes in growth will be detected over a period of time shorter than one school year.

The target outcome of the performance feedback in this study is total words written (TWW). TWW was chosen because it is a basic measure of writing fluency and is easily understood by elementary-aged students. It is possible that some students, when working to increase the amount that they write, sacrifice the quality of their writing. Therefore, it was determined that a CBM-WE fluency metric incorporating some evaluation of quality should be included as a secondary dependent variable. Reviews of the technical adequacy of CBM-WE suggest that choice of metric to use depends on students' grade level and the purpose of the measurement (McMaster & Campbell, 2008; McMaster & Espin, 2007). Although CWS, %CWS, and CIWS all demonstrate high criterion validity for elementary-aged students, not all metrics detect growth in fluency. For example, one more recent study failed to detect growth from fall to spring for all three metrics: CWS, %CWS, and CIWS (McMaster & Campbell, 2008). A previous study demonstrated growth in CWS and CIWS with a somewhat larger effect for CWS (Malecki & Jewell, 2003). CWS is highly correlated with TWW and both are fluency measures; therefore, CWS was chosen as the secondary dependent variable to be most likely to detect growth.

Given the recommendations in the area of writing fluency (Abbott & Berninger, 1993; Eckert et al., 2006, 2008; Graham & Harris, 1997), it was predicted that performance feedback on TWW (performance feedback condition) would have the greatest impact on CBM-WE and that weekly practice (practice-only condition) would have a greater impact on CBM-WE than typical instructional practices (instructional control condition). A second aim of this study was to examine and describe the variability in writing fluency associated with two key variables, sex and initial level of writing fluency. Because research has demonstrated that girls generally outperform boys in assessments of writing fluency (Berninger & Fuller, 1992; Camarata & Woodcock, 2006; Fearrington et al., 2014; Kim et al., 2014; Malecki & Jewell, 2003; Martin & Hoover, 1987), it was hypothesized that girls would have a higher writing fluency level estimate than boys. Differences in growth estimates between girls and boys were explored given an initial finding (Fearrington et al., 2014) of a significant interaction between sex and time in grade 3. It was also hypothesized that students with a lower level of initial writing fluency (i.e., below the 25th percentile) would have greater growth than students initially writing at a higher level (i.e., above the 25th percentile). This hypothesis was based on research suggesting that lower-performing students demonstrate greater improvement in writing fluency than groups of students who begin at a higher rate (Shinn et al., 1982).

A final aim of this study was to examine students' acceptability ratings of the procedures used in the study (i.e., receiving feedback, and writing in response to CBM-WE probes). The importance of examining the acceptability of classroom-based procedures has been advocated for many years (Eckert & Hintze, 2000; Elliott, Witt, & Kratochwill, 1991), and it continues to be identified as an important area of inquiry that is infrequently assessed (National Association of School Psychologists, 2005; Stebbe Rowe, 2012). In addition to providing support for the social or ecological validity of the procedures (Wolf, 1976), obtaining subjective evaluations of procedural acceptability provides some evidence regarding the cultural appropriateness of the procedures for changing the students' academic skills (Kratochwill & Stoiber, 2000, 2002). Because no previous research examined students' perceptions of the practices used in this study, the analyses were considered exploratory and no a priori hypotheses were developed.

METHOD

Participants and setting

The demographic data for the sample of 133 third-grade students participating in the study is described in Table 1. Although 20 students (15%) were receiving special education services, none were classified as learning disabled in writing. Chi-square statistics revealed no significant differences across the three conditions for sex, χ^2 (2, N = 133) = 1.71, p = .43, ethnicity, χ^2 (6, N = 133) = 8.36, p = .21; age, F (2, 132) = 1.71, p = .19; or special education status, χ^2 (2, N = 133) = 2.43, p = .30.

The setting for the study was nine general education classrooms housed in three elementary schools in a mid-sized city in the northeast. Each school enrolled students in kindergarten through fifth grade. These schools were considered to represent an underserved population as an average of 78% (range = 66% to 86%) of the students attending the three schools qualified for free or reduced-price lunches, and over 80% of the students represented underserved minority groups.

Materials

Test of Written Language — Third Edition—The Test of Written Language — Third Edition (TOWL-3; Hammill & Larsen, 1996) is a standardized, norm-referenced assessment of written expression skills for children, aged 6 through 16. In this study, the Spontaneous Writing subtest was administered for the purpose of quantifying each student's general writing abilities. The Spontaneous Writing subtest requires students to look at a picture, plan a story, and write a story for 15 min. The story is evaluated in three areas: (a) contextual conventions (e.g., punctuation, capitalization, and spelling), (b) contextual language (e.g., vocabulary and sophistication of sentence construction), and (c) story construction (e.g., plot, character development, and style). The manual reports high (i.e., greater than .80) internal consistency, interscorer, and parallel-form reliability coefficients for the resulting Spontaneous Writing Quotient for 8- and 9-year-old children. In terms of criterion-related validity, the Spontaneous Writing Quotient demonstrated a moderate association (r = .50) with the writing score from the Comprehensive Scales of Student Abilities (Hammill & Hresko, 1994).

Curriculum-based measurement probes in written expression—CBM-WE probes were administered over the course of the study. Similar to current standard practice, each weekly probe contained a different story starter, for a total of 8 different story starters used across the study. All conditions were administered the same story starter each week. Story starters are short sentence fragments that provide an idea to the students for writing a narrative story (e.g., "I was talking to my friends when all of a sudden …"). Current standard practice is to use story starters as parallel forms despite lack of empirical evidence for that practice. The story starters chosen for this study were identified as appropriate for use with diverse populations (McMaster & Campbell, 2006); however, they were not thoroughly evaluated for use as parallel forms. Pilot use of these story starters indicated low moderate to high correlations between each time point when examining TWW (r = .29 to .81) and CWS (r = .37 to .76; Truckenmiller, 2007). A slightly higher range of correlations were found in the current study for TWW (r = .52 to .83) and for CWS (r = .53 to .79), which is also similar to the range in published sets of CBM-WE story starters (i.e., AIMSweb; Powell-Smith & Shinn, 2004).

In this study, each written response to the CBM-WE probes was evaluated for two metrics: TWW and CWS. TWW was calculated by counting every grouping of letters separated by a space. Words were counted regardless of spelling or accuracy; however, numerals were not counted. Based on scoring procedures outlined by Shapiro (2004), CWS was computed by analyzing each adjacent word for correct punctuation, capitalization, spelling, and syntax. A CWS was counted when two adjacent words were correctly-spelled and the words followed a syntactically and grammatically-allowable sequence. A CWS was also marked and counted between a word and a punctuation mark, a punctuation mark and a word, and between two punctuation marks if the writer did not violate any of the rules for punctuation. A detailed copy of the rules used for scoring CWS is available from the first author by request.

TWW was chosen as it provides a highly reliable measure of fluency that is most commonly used with elementary-aged children and CWS provides a more precise measure of fluency as well as an indication of quality of written work. These two metrics, along with WSC, are the most commonly used to assess elementary-aged children's written fluency skills (Espin et al., 2000). The technical adequacy of these metrics of CBM-WE has been explored in 15 studies of elementary-aged children with comprehensive reviews provided in the AIMSweb manual (Powell-Smith & Shinn, 2004) as well as a meta-analysis (McMaster & Espin, 2007). Overall, the reported parallel form reliability coefficients for TWW in studies that include grade three are moderate to high (range, r = .51 to .99), as well as interscorer agreement (range, 91% to 99%). In the few studies that examined the psychometric properties of CWS in third grade, parallel form reliability (r = .46) and interscorer agreement (range, 86% to 98%) was lower than the estimates for TWW or WSC. However, studies examining the validity of the three metrics indicate that the CWS metric was more highly correlated (range, r = .31 to .85) with criterion measures than either TWW (range, r = .08 to .82) or WSC (range, r = .21 to .88; McMaster & Espin, 2007; Powell-Smith & Shinn, 2004).

Procedural acceptability assessment packet—A brief survey was administered to all students to assess their perceptions of the writing procedures used in the study. This assessment included a series of six questions regarding the students' attitudes toward the procedures used, as well as their own writing skills. A 5-point Likert-type response system was used with responses ranging from not at all to very, very much. Students assigned to the performance feedback condition were asked to respond to two additional questions regarding the students' perceptions toward receiving feedback. The questions contained in the packet were used in order to obtain a short descriptive evaluation of the students' perceptions on the specific procedures used during this study.

Curriculum-based measurement in mathematics—Curriculum-based measurement probes in mathematics (CBM-M) were administered over the course of the study to participants assigned to the instructional control condition. Each probe contained 60 mixed basic addition and subtraction facts, using numerals no higher than 12. Mixed addition and subtraction were targeted as third grade students are expected to be proficient (fluent) with the addition and subtraction of whole numbers by the end of third grade (National Mathematics Advisory Panel, 2008). The psychometric properties of CBM-M have been examined in a number of studies. Interscorer agreement for the number of digits correctly computed is high (range = 83% to 93%; Shinn, 2004). All conventional reliability estimates (i.e., internal consistency, test–retest, and parallel forms) were higher than .70, with stronger reliability for single-skill CBM-M probes than mixed-skill CBM-M probes (Foegen, Jiban, & Deno, 2007; Shinn, 2004). Validity coefficients in a review of 17 studies were reported to be modest, ranging from .50 to .70, but similar to other tests of mathematics achievement (Foegen et al., 2007).

Procedures

Teachers and students attending the three elementary schools were invited to participate in the study. Once teacher consent was obtained (9/9, 100% consent rate), parental consent was obtained for 147 out of the 199 students (74%). Of the students whose parents consented, 100% provided student assent to participate in the study. Students were then screened for eligibility, which included (a) not experiencing severe motor deficits that preclude students from composing written stories; (b) not experiencing significant cognitive deficits; (c) speaking English as their primary language; (d) not classified as learning disabled in writing; (e) not receiving one-to-one instruction or having a Section 504 plan indicating additional instructional modification; (f) demonstrating minimum proficiency by writing at least seven words on the baseline CBM-WE probe; and (g) legibly scribing 90% of 10 alphabetic letters dictated to them. A total of eight students (5%) were excluded from the study because they did not meet the inclusionary criterion of writing seven or more words on baseline CBM-WE probe.

Following the eligibility screening, 139 participating students were randomly assigned to one of three conditions: (a) a performance feedback condition, (b) a practice-only condition, or (c) an instructional control condition. Two students were unable to participate in the conditions due to schedule conflicts, and four students were excluded from the analyses due to limited data (i.e., missing baseline assessment). This resulted in a final sample of 133 third-grade students participating in the performance feedback condition (n = 46), the practice-only condition (n = 39), or the instructional control condition (n = 48). All eligible students participated at the same time. Each condition was administered in a group format in the students' regular classrooms. In order to accommodate the random assignment of students to condition, two thirds of the students had to transition to a different classroom. Students in the performance feedback condition were called the green group, the practice-only condition was the gold group, and the instructional control condition was the blue group. During the weekly scheduled time, all of the students in third grade would move classrooms so that all of the students in the green group at that school were in one classroom, all students in the gold group were in another classroom, and all students in the blue group were in the third classroom. Sessions were conducted simultaneously with two research assistants in each classroom administering the conditions. Teachers also remained present in the classrooms during sessions. Data collection began in the winter of the school year and occurred weekly for nine weeks. All three schools participated during the same time frame.

Eligibility and baseline assessment phase—To determine students' eligibility to participate in the study based on handwriting legibility, participants were asked to print 10 lowercase letters from the alphabet. Subsequently, Form A of the Spontaneous Writing subtest of the TOWL-3 (Hammill & Larsen, 1996) was administered in a group format to determine overall writing ability at baseline. During the second baseline assessment session, two CBM-M probes were administered followed by one CBM-WE probe to all eligible students. The results from the CBM-WE probe were used at the next session (i.e., the first session of the intervention phase) to provide feedback to the students assigned to the performance feedback condition. The results from the two CBM-M probes were averaged and used to provide feedback to the students assigned to the instructional control condition.

Individualized performance feedback condition—Each week, students in the performance feedback condition received a packet that contained individualized performance feedback from the previous week and a new CBM-WE probe to respond to. Following a script read to the entire group of students, the research assistant explained how to read the feedback page of the packet and subsequently lead the students through completion of the next CBM-WE probe. In the oral description, the research assistant explained that the total number of words written was computed by counting all words that each student wrote. The visual portion of the feedback (i.e., the feedback page) was individualized for each student. At session 2, the feedback page included a box containing the total number of words the student wrote at baseline. For each of the subsequent six sessions, the feedback page also displayed an arrow either pointing upward or downward to indicate an increase or decrease, respectively, of the total number of words the student wrote (i.e., TWW). Students were then provided 1 min of quiet story planning, and 3 min of continuous writing within the context of the CBM-WE probe.

Practice-only condition—Procedures identical to those described for the performance feedback condition were followed concurrently, excluding the individualized feedback page of the packet. The story starters in the CBM-WE probe were identical to the story starters used in the performance feedback condition. Procedures for this condition included only the 1-minute planning and 3-minute writing time and were therefore finished prior to the other two conditions. Students in this classroom worked on a word search until the other two classrooms were finished so that students could all transition back to their regular classrooms at the same time.

Instructional control condition—The instructional control condition was designed to closely resemble the performance feedback in writing condition by providing a similar instructional experience as the performance feedback but focused on an unrelated academic skill: mathematics computation. Although this condition was considered to be a control condition, students may have benefited from the feedback in basic mathematics computation. During these sessions, the students were given two CBM-M probes targeting basic addition and subtraction facts (spanning 2 min each). Prior to responding to the CBM-

M probes, individualized performance feedback on digits correct was provided to each student, similar to the performance feedback condition. Because students were given two CBM-M probes at each session, the digits correct feedback was calculated as the average number of digits correct on the two probes. On three occasions (session 1 or baseline, session 4, and session 8) students assigned to the instructional control condition were administered a CBM-WE probe identical to that previously described for the practice-only condition. The story starters were identical to those used by the feedback and practice-only conditions at sessions 1, 4, and 8.

Procedural integrity

A total of four doctoral students in school psychology, one university faculty member in school psychology, and four advanced undergraduate psychology majors administered the experimental conditions. All assistants practiced delivering each condition following the script. Procedural integrity was assessed in two ways: a permanent product aligned with the procedural script obtained at each session and a secondary observation conducted in 73% of the sessions. Overall, procedural integrity for each condition was high. In the performance feedback condition, the primary research assistant administered an average of 99% of the steps (range = 96% to 100%). Observations by the secondary research assistant occurred during 67% of the sessions confirmed that an average of 99% of steps were administered as scripted (range = 96% to 100%). In the practice-only condition, the primary research assistant administered an average of 99% of the steps (range = 89% to 100%). Observations by the secondary research assistant occurred during 81% of the sessions confirmed that an average of 99% of steps were administered as scripted (range = 89% to 100%). In the instructional control condition, the primary research assistant administered an average of 99.8% of the steps (range = 95% to 100%). Observations by the secondary research assistant during 70% of the sessions confirmed that an average of 99.8% of steps was administered as scripted (range = 95% to 100%).

Interscorer agreement

Ten advanced undergraduate psychology majors were trained to use a CBM-WE scoring manual to score TWW and CWS, and they practiced scoring prior to the beginning of the study. In order to begin scoring, research assistants had to reach 100% agreement on one CBM-WE probe. A total of 40% of the CBM-WE probes were randomly selected and re-scored for TWW and CWS. On average, interscorer agreement was high and similar to previously published results. The average interscorer agreement for TWW was 99% (range = 71% to 100%) and was calculated by dividing the lower total count with the higher total count and multiplying by 100%. The average interscorer agreement for CWS was 95% (range = 75% to 100%) and was calculated as number of agreements divided by agreements plus disagreements.

Data analysis

Multi-level growth modeling was used in this study to account for the nested nature of academic skill growth within students (Burchinal, Nelson, & Poe, 2006; Raudenbush, 2001; Singer & Willett, 2003), and it matches the goals of curriculum-based measurement in four important ways: (a) it retains individual differences by modeling each individual’s intercept and slope estimates, (b) it is sensitive to detecting variables affecting incremental changes in outcome measures, and (c) it accommodates missing data, and (d) it can detect predictor variables that may affect outcome measures. Data were inspected for violations of normality (e.g., skew) by reviewing (a) histogram distributions of residual error values, (b) quantile–quantile plots, (c) percentile plots, and (d) tests of normality. No significant deviations from normality were found. Multilevel modeling was chosen for its amenability to evaluate growth over time with missing data points. All students included in the analyses participated in at least 50% of the sessions (producing 4 data points). There was a total of 95 missing data points out of 824 possible data points (12%) with the instructional control condition, the practice-only condition, and the performance feedback condition missing 8%, 16%, and 9% of data points, respectively.

Multilevel modeling (PROC MIXED in SAS V9.1, 2004) with restricted maximum likelihood estimation procedures was used in the presence of these missing data (Peugh, 2010). Models were tested to assess the differences in growth of students' writing fluency across the three instructional practices and adding sex and initial level of writing fluency as predictors. A three-level model was specified with assessments (weekly session 1 through session 8) nested within individual students (N = 133) and students nested within schools (N = 3). The time variable (i.e., weekly session) was centered on the final observation (session 8) in order to directly test the effects of feedback on students' writing fluency in a brief time period. An alpha level of .05 was chosen for all significance testing. Given that there are two dependent variables (TWW and CWS) in the study and the Benjamini–Hochberg correction is recommended for use in experimental research in education (Schochet, 2008), the acceptable p value for significance tests regarding TWW and CWS in this study is .041 (Benjamini & Hochberg, 1995).

Predictor variables, sex and initial level of fluency, were coded as categorical variables. Following the AIMSweb winter benchmark normative table (Pearson Education, Inc., 2009), students' initial level of fluency in the current study was categorized as at-risk (below the 25th percentile or 23 TWW or less) or not at-risk (above the 25th percentile or 24 TWW or greater). For the statistical model predicting CWS, students were categorized as frustrational (below the 25th percentile or 14 CWS or less) or instructional (above the 25th percentile or 15 TWW or greater). The same modeling procedures were followed separately for each dependent variable: TWW and CWS.

RESULTS

Descriptive analyses

In order to further describe the sample of participants, performance on baseline measures of the participating students' writing abilities are reported in Table 2. Similar to other studies of writing, the description of these students' initial writing abilities depends on the writing assessment used. Average performance on the TOWL-3 Spontaneous Writing Quotient (M = 84.31, SD = 32.05) fell more than one standard deviation below the mean. However, the results of the CBM-WE probe indicated that students wrote an average of 31.50 TWW in 3 min (SD = 12.44) and 26.86 CWS in 3 min (SD = 12.22). The students' baseline performance in this sample corresponded to the 50th percentile for third-grade students

(Pearson Education, Inc., 2009). In an unrelated academic skill, basic mathematics fact fluency, the mean performance of all students (M = 28.91, SD = 13.09) fell between the 50th and 75th percentiles for third-grade students (Pearson Education, Inc., 2009). Students' performance on each of these baseline measures (TOWL-3 Spontaneous Writing Quotient, CBM-WE, and CBM-M) was significantly correlated with one another (range $r = .33$ to .91).

One significant difference in performance on CBM-WE was found between boys and girls with the girls significantly outperforming the boys in TWW, $F(1,131) = 5.15$, $p = .025$ and CWS, $F(1,131) = 5.85$, $p = .017$. Any differences based on school effects were leveraged by having random assignment of students to conditions stratified by school and nesting students within schools in the multilevel model. Baseline differences between the three conditions were not found on the primary dependent variables (i.e., TWW & CWS), indicating that randomization had been successful.

Total words written

The fixed and random effects for the unconditional model and models examining the hypothesized variables are summarized for TWW in Table 3. The preliminary model estimated variances for the intercept and slope coefficients for students within classrooms and students within schools; however, the results indicated that no significant variance was present at the classroom or school level. Because the students were randomized to condition at the school level instead of the classroom level, school remained the upper level nesting variable. Subsequently, the random effects for the upper level were fixed, leaving only the student level variances freed for estimation. Results from this model indicated that 73% of the total variance in scores was due to between-student differences at the end of year in TWW, whereas b1% was due to slope. The initial growth model indicated that there was significant growth over time for TWW for all students, $t(595) = 3.37$, $p < .001$, gaining an average of 0.51 TWW each week and reaching an estimated average total of 34.59 TWW by the end of the study.

When instructional effects were entered into the model (see Table 3, third column), the control students were observed to have correctly written 32.15 total words at the final assessment of the study ($p < .001$). These students, on average increased their writing skill by 0.37 words per week; however, this estimated trajectory was not found to be statistically significant. Students in the performance feedback condition achieved a higher level of TWW at the end of the study compared to the control by an average of nearly 10 words ($p < .001$). This magnitude of difference in performance at the final assessment corresponded to a calculated Hedges' g of 0.66, which represents a medium effect (Cohen, 1988). This difference also demonstrates clinical significance as the separation between the two groups is the same distance as between a national grade three mean and the grade four mean (Pearson Education, Inc., 2012). The growth rate for those in the performance feedback condition was also stronger than the control; gains over time were 0.93 words per week greater than the control ($p = .007$).

The comparison between the practice-only group and the instructional control group did not yield any statistically significant differences; indeed, an examination of the estimated growth for the practice condition indicated that their growth actually decreased over time by approximately 0.70 total words per week, and was differentiated from the control group (p = .03). A final comparison of the TWW outcome was a contrast test for differences between the feedback and practice groups. Results indicated that the feedback participants outperformed their practice counterparts at the final assessment by an estimated 13.09 TWW, t(130) = 4.22, p < .001, and their growth in TWW was differentiated by 1.63, t (593) = 5.06, p < .001. The difference between the feedback group and the practice group represented a moderate effect (Hedges' g = 0.65; Cohen, 1988). The inclusion of instructional conditions in the model accounted for 67.7% of the random variance in growth in TWW as well as 17% of the final status variance. The model estimates were used to show the predicted growth trajectories of each group in Fig. 1.

The final TWW model also included sex effects and resulted in the best model fit based on goodness of fit statistics, as well as explaining an additional 4.8% of the variance in growth in TWW and an additional 3.7% of variance in final level of TWW. Girls' level of fluency as measured by TWW was significantly greater than boys at baseline and remained at a greater level through the end of the study, t(127) = 2.82, p = .006. This result is consistent with the second hypothesis as girls wrote an average of 5.85 more TWW than boys at the end of the study. Exploratory analyses of the differences between boys' and girls' rate of growth in writing fluency across a short period of time indicated no statistically significant differences in growth for TWW, t(590) = 1.70, p = .09. Furthermore, the final hypothesized predictor variable, initial level of fluency – characterized as at-risk according to AIMSweb normative data (Pearson Education, Inc., 2009) – did not differentially predict variation in growth in TWW, t(592) = 2.29, p = .13. This finding suggests that participants demonstrated the same rate of growth over the course of the study, regardless of risk categorization at baseline.

Correct writing sequences

The fixed and random effects for the models examining the hypothesized variables are summarized in Table 4 for CWS. The initial growth model indicated that there was significant growth over time for CWS, t(595) = 3.58, p < .001. Students gained an average of 0.46 CWS each week and reached an average total of 29.58 CWS. It should be noted that this growth model and subsequent models of CWS were specified differently than the TWW models due to constraints in estimating the random variance in slope. This encounter with boundary constraints has been reported in other research studies investigating writing outcomes (Coker, 2006). Therefore, only the CWS models (with estimates reported in Table 4) follow Singer and Willett's (2003) recommendation to simplify the model through "fixing" the slopes by removing the random effects of time from the model (p. 153).

Although CWS was not targeted by the feedback intervention, the general pattern of results for CWS was similar to TWW. The students in the performance feedback condition ended at a higher level than the control group, t(130) = 2.07, p = .041. However, this result was not statistically significant after applying the Benjamini–Hochberg correction (Benjamini & Hochberg, 1995). In order to further describe the relationship between the performance feedback condition and the control group, Hedges' g was computed to be 0.34. This represents a small to moderate effect which may be clinically significant to some educators considering that the growth of students across a full school year with typical instructional practices yields a small effect size (Malecki & Jewell, 2003).

The students in the practice condition did not perform significantly better than the control group, $t(130) = -1.27$, $p = .21$. The estimated means of final level of writing fluency were 33.82 CWS for the feedback group, 28.57 CWS for the control group, and 25.18 CWS for the practice-only group. When examining growth in CWS, differences between the performance feedback group and the control group approached statistical significance, $t(593) = 1.93$, $p = .054$, but did not exceed the a priori alpha threshold. The difference in growth between the practice-only condition and control condition also approached statistical significance and favored the control condition, $t(593) = -1.95$, $p = .052$, but did not exceed the a prior alpha threshold. On average, students in the performance feedback condition gained an estimated 1.00 CWS each week, students in the practice-only condition lost an estimated mean of 0.22 CWS per week, and students in the instructional control condition gained an estimated mean of 0.41 CWS each week.

Exploration of possible predictor variables (i.e., sex and initial level of fluency) was pursued in order to further explain the nature of growth in CWS over an 8-week period. Similar to the results for TWW, girls and boys did not have statistically significantly different rates of growth, $t(590) = 1.07$, $p = .28$. However, sex explained 4.0% of the random variance in level of CWS over and above the variance explained by experimental condition. Furthermore, the model that included experimental condition and sex as predictors had the best model fit (see Table 4, far right column). The estimated difference between girls and boys final level of performance was 7.22 CWS and approached significance, $t(127) = 1.93$, $p = .06$. Students' categorization as 'at-risk' based on their initial level of CWS did not differentially predict growth in CWS over the course of the study, $t(592) = 0.96$, $p = .33$.

Procedural acceptability

Table 5 provides the mean rating (on Likert-type scale of 1 to 5 with 5 indicating high acceptability) of each item of the procedural acceptability assessment as rated by the student participants at the end of the study (i.e., the last session of the study). The six items contained on the scale demonstrated moderate levels of internal consistency ($\alpha = .70$). A statistically significant difference in acceptability of the story-writing procedures was found between the three conditions, $F(2,104) = 3.36$, $p = .038$. On average, students in the practice-only condition rated the story-writing procedures the highest ($M = 3.94$; $SD = 0.88$), followed by students in the performance feedback condition ($M = 3.71$; $SD = 0.95$) and then students in the instructional control condition ($M = 3.38$; $SD = 0.83$). Acceptance of the procedures used in this study by students who responded to CBM-WE probes each week was generally moderate. These ratings are somewhat lower than previously demonstrated student acceptability of procedures (Eckert et al., 2009). Additionally, students in the performance feedback condition were asked to respond to two questions assessing their acceptability of the feedback procedures. Students assigned to the performance feedback group found the feedback procedures to be highly acceptable, with mean ratings greater than 4.25 on a 5-point Likert-type scale.

DISCUSSION

Building upon previous positive findings in quasi-experimental studies for the use of performance feedback to improve writing fluency (Eckert et al., 2006, 2008), this study employs empirical methods (e.g., control condition; random assignment of students to condition; controlling for school level effects) supporting performance feedback as an effective component of writing intervention. Not only did performance feedback improve writing fluency at a statistically significant level with a moderate effect size, but it also demonstrated clinical significance. In just 7 weeks, the students assigned to the feedback condition in the current study reached 41.6 TWW, a level of writing fluency that exceeds the designation for instructional level for third grade (i.e., 37 TWW in 3 min; Mirkin et al., 1981). This finding is particularly important as none of the group means based on other variables (e.g., girls and School A) were at the instructional level at baseline (see Table 2). Within the context of instructional decision-making, a shift in mean writing fluency from below instructional level to above instructional level is a highly desirable outcome. The curricular leadership team in a school will often make decisions about the effectiveness of particular instructional practices based on data like this (Fuchs & Fuch, 2006).

On average, the students in the performance feedback condition in the current study gained more than twice the TWW and CWS per week as the national average of 0.4 (Pearson Education, Inc., 2009). The potential added benefit is that the feedback intervention targeting TWW also had a small residual impact on a measure of writing fluency quality, CWS. The similar model estimates between TWW and CWS may be due to the high correlation between TWW and CWS that was found in this study and previous studies. Given the similarities and differences in results between the two metrics, many questions exist for future research into the nature of the relation between a writing metric that is solely based on fluency (TWW) and a writing metric that incorporates some element of writing quality (e.g., spelling, punctuation, and grammar).

The limited impact of the practice-only condition in this study contributes to mixed evidence regarding the amount of growth that occurs in response to practice with CBM-WE probes. The practice-only groups in previous studies demonstrated noticeable growth over time, about 1 TWW per week (Eckert et al., 2006, 2008). Again, it is most likely that the differences from previous studies are due to the inclusion of significant control variables (e.g., instructional control condition and school-level variables) in the current study. This new information suggests that feedback should be coupled with practice with CBM-WE in order to produce significant growth in writing fluency. Replication will be essential to support this conclusion; however, practitioners should be aware that effects may not be obtained by simply using repeated practice for groups of students. For example, in the area of reading, repeated practice with a passage produces significant positive effects on reading fluency growth (Daly, Martens, Hamler, Dool, & Eckert, 1999). The current study suggests that the same may not be true for writing fluency. Although, the effects of more intensive repeated practice with CBM-WE have not been explored.

Despite the limited research base regarding the acceptability of academic interventions (Eckert & Hintze, 2000), the present investigation supplements our understanding of this construct related to students in the content area of written expression. The present investigation examined students' acceptability of the CBM-WE probe within three different instructional practices. Higher levels of acceptability were endorsed by students in the practice and performance feedback conditions, but students in all three instructional conditions rated CBM-WE favorably. This finding is useful for school-based practitioners to consider when recommending assessment and instructional practices in the classroom. Students who were exposed to the writing procedures on a more frequent basis (i.e., in the performance feedback and practice only conditions) seemed to have higher levels of acceptability. In addition, these findings further our understanding of evidence-based academic interventions that are viewed as fair and reasonable by their respective consumers (i.e., the students participating in the interventions).

Growth in students' writing fluency

One critical consideration and one of the identified barriers to intervening with elementary-aged students' writing fluency are the dearth of knowledge about the expected amount of growth in response to CBM-WE (Eckert et al., 2009). In general, reports indicate that elementary-aged children experience very limited writing fluency growth over a school year (i.e., 8 TWW or less; Marston et al., 1981; McMaster & Campbell, 2006). AIMSweb, the only source of current national norms for CBM-WE metrics, reports that the median rate of growth for third-, fourth-, and fifth-grade students is about 0.4 TWW and 0.4 CWS per week (Pearson Education, Inc., 2009). However, these reported estimates are based on changes in the mean performance of a large sample from fall to winter to spring. This type of calculation fails to represent true growth because it examines mean differences as opposed to growth patterns within individual students (Cronbach, 1976). The results of the current study indicate that the rate of growth for third-grade students (receiving typical instructional practices) over the course of 8 weeks was not significantly greater than zero. Although estimates of growth provided by the current study are limited by a relatively small sample size, the results do provide additional evidence that late elementary-aged students gain approximately 0.4 words and correct sequences per week (Pearson Education, Inc., 2009). Specifically, students assigned to the instructional control condition in this study gained approximately 0.37 TWW and 0.41 CWS per week.

Some caution is warranted when interpreting the growth estimates for the instructional control group in this study. Given these students' unexpected rate of growth in writing fluency, it is possible that other variables within the study affected their growth in writing fluency. The students in the instructional control group were given performance feedback on their mathematical calculation fluency during three sessions between the baseline and midpoint CBM-WE measurement and during two sessions between the midpoint and last CBM-WE measurement. It is possible that the students in the instructional control condition generalized their performance from the CBM-M probes on which they were receiving performance feedback to the CBM-WE probes. It can be hypothesized that the similarity of instructions for the CBM-WE and CBM-M tasks and the presence of a stopwatch cued the students in the instructional control condition to perform more quickly on the CBM-WE probes based on their previous experience of feedback and a time limit on the CBM-M probes. In fact, students in the instructional control condition reported similar acceptability of being timed while writing as the students in the writing feedback group, which was considerably less than the acceptability reported by the students in the practice-only group. Therefore, it is likely that the students in the control condition generalized their perception of the performance expectations to the writing task.

Variables affecting students' writing fluency: sex

The current study suggests that boys' growth in writing fluency is neither greater nor less than girls' growth in writing fluency over an 8-week period but represents an area for further exploration. The girls' significantly higher performance in this study corroborate previous research studies demonstrating gender differences in writing performance across the lifespan (Berninger & Fuller, 1992; Berninger et al., 2008; Fearrington et al., 2014; Kim et al., 2014; Malecki & Jewell, 2003; Marston et al., 1981; Persky et al., 2003). The current study did not find that boys and girls had significantly different rates of growth; it contributes to a very small literature base (Fearrington et al., 2014) suggesting mixed evidence of sex differences in writing fluency growth. If future research continues to find other language and cognitive abilities that also have sex differences have an impact on writing outcomes, then practitioners may expect to identify larger numbers of boys needing additional instruction and intervention with writing fluency than girls. Kim et al. (2014) examined quite a few language and cognitive abilities related to literacy and found that they explained approximately one third of the sex differences in writing fluency. They suggest that other cognitive abilities may need to be explored, but boys' and girls' motivation with writing may also be important. Instructional implications will depend on identifying those factors that are contributing to growth in writing fluency and exploring the causes of gender differences may be a way to facilitate identification of those areas that affect writing and girls based on the factors affecting their performance. The findings in the current study demonstrate that girls did not grow faster and that boys did not close the gap during a brief period of performance feedback.

Variables affecting students' writing fluency: instructional level

Results failed to find that categorization of elementary-aged students' level of fluency differentially predicted students' growth trajectory in writing fluency. It is not possible to determine if one instructional practice (feedback, practice, or typical instructional practices) is differentially beneficial for students identified as performing at a frustrational level in writing fluency due to a lack of statistical power in the current study. The finding that classification of instructional level was not a significant predictor of writing growth suggests that a designation of frustrational level is not needed to effect growth in writing fluency. However, the lack of findings in this area may be due to statistical restrictions from confining students to two groups: at-risk and not at-risk. This designation was used due to its practical significance. A much larger study in the future may be able to provide differential growth rates based on students' initial level of writing fluency similar to the differential growth rates provided for curriculum-based measurement in reading (Deno, Fuchs, Marston, & Shin, 2001).

Although this study examined two predictor variables, sex and instructional level, other variables may be contributing significantly to writing fluency growth. For example, Coker

(2006) found that school-level variables (i.e., first-grade teacher, classroom literacy environment) differentially predicted writing growth across a 3-year period. Furthermore, the instructional environment for writing may be difficult to characterize without including an observation of the writing instruction occurring in each classroom. As Cutler and Graham (2008) demonstrated, teachers' writing instruction as well as materials and philosophies used to guide writing instruction are highly variable. It is possible that variables related to writing outside of the parameters of this study are contributing to variation in writing fluency growth.

Limitations

The most concerning limitation of this study is the difficulty in determining the amount of variability that is due to differences in student background knowledge, motivation, and difficulty of the story starter stimuli. Consideration of the various ways difficulty in story starters may affect student performance were considered throughout the design of the study and analysis of the data. The story starter stimuli used in this study were chosen because they are the only set of story starters that has been evaluated as appropriate for a wide variety of cultures and potentially limited background knowledge (McMaster & Campbell, 2006). However, the size of the effect due to difficulty level of story starters has not been previously published, so it is difficult to determine the possible effect of story starter difficulty on the current study. When studying curriculum based measurement in reading, others (e.g., Christ, White, Ardoin, & Eckert, 2013; Francis et al., 2008) have explicitly tested for form effects using a counterbalanced administration design. Because the story starter stimuli in CBM-WE can elicit a wide range of possible responses, it is not as bound to the stimulus as curriculum-based measurement in reading, where there is only one correct response for each word in the stimulus and it is much more sensitive to variability in difficulty. Powell-Smith and Shinn (2004) reported that parallel form reliability coefficients for CBM-WE are moderate to high (range = .51 to .99). Furthermore, the story starters in the current study could not be counterbalanced in order to maintain administration similarity across the three conditions. Order effects in the current study may have been mitigated by including the practice-only condition. Since the practice condition did not have significant growth over the eight weeks of the study, we can determine that either (a) practice effects with CBM-WE over the course of eight weeks is not significantly different from zero, or (b) the variation in difficulty of these story starter stimuli may have counteracted any detectable practice effects. Future study on the story starters with a counterbalanced design (similar to the study conducted by Francis and colleagues) is needed to determine the magnitude of the impact of story starters on students' performance.

There were also notable limitations associated with the experimental design. Although the study hypotheses were not revealed to the experimenters, the experimenters were not blind to the participants' condition assignment for administration or for scoring of CBM-WE probes. Procedural integrity and interscorer reliability were monitored in order to capture blatant instances of experimenter bias, but it is plausible that some form of bias was introduced. Further, in order to have a true control group, measurement using CBMWE could only occur on three occasions throughout the study, which resulted in an unbalanced number of data points across the three conditions (eight waves of data for the feedback and practice-only condition and three waves of data for the control condition). The unbalanced number of data points likely caused the CWS model to reach boundary constraints and prevented us from allowing the slopes to vary randomly. Furthermore, by having only three data points, performance by the control group was constrained to a linear trajectory, and growth between the three data points could only be modeled in a linear fashion. The likelihood that writing growth, particularly when examined over a short measurement period, occurs in a nonlinear fashion is unlikely. To date, a nonlinear model of writing growth for elementary-aged children's writing fluency has not been found (Coker, 2006; Eckert et al., 2008; Marston et al., 1981).

It is also possible that the instructional control condition did not represent a true control group. Students in the instructional control condition received performance feedback on math probes in order to closely approximate the performance feedback in writing condition but target an unrelated skill: mathematical calculation fluency. However, the performance feedback provided for mathematical calculation fluency in the instructional control condition may have also unintentionally impacted the instructional control students' writing fluency. This possibility limits conclusions about the effect of practice with CBM-WE and the interpretation of writing fluency growth estimates in the absence of intervention. The growth rate of students in the instructional control condition sample may have been higher than the actual population and the sample's growth rates in the practice-only condition may have been suppressed.

In addition, the standard error of the slope estimates for the practice-only condition (0.33) and the instructional control condition (0.24) were almost as large as the respective slope estimates. This finding draws into question the reliability of the slope estimates. To date, research has not examined the standard errors of the slope estimates for intervention studies in writing. In the content area of reading, it has been demonstrated that the standard errors of the slope estimates tend to be large and less stable over shorter periods of time (i.e., less than 10 weeks; Christ, 2006; Christ, Zopluoglu, Monaghen, & Van Norman, 2013), which negatively impacts the reliability of making educational decisions at the individual level and idiographic hypothesis testing (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013; Christ, White, et al., 2013; Christ, Zopluoglu, et al., 2013). As a result, the confidence intervals associated with the slope estimates should be heavily relied upon when estimating students' writing growth over time. Moreover, it is possible that the large standard error of the slope estimates was influenced by other unaccounted or uncontrolled variables that affected the students' rate of writing growth.

Although the use of multilevel modeling can improve parameter estimates by accounting for group-level variation to a greater extent than a conventional linear model (Raudenbush, 2001) as well as accounting for the nested structure associated with school-based research, a number of within-subjects or within-school variables were not taken into account. For example, school assignment was accounted for in the present study, but additional variables, such as student ethnicity and race, teacher experience, quality of the reading and language arts curriculum, and school climate, were not. Given that it is common for school-based research to include multiple uncontrolled and interacting variables, it has been recommended (e.g., Daly, Hintze, & Hamler, 2000; Singer & Willett, 2003) that reliably known and measured variables be relied upon when specifying multilevel models. As a result, further exploration of these alternative variables was not prioritized in the current study.

A final consideration regarding external validity is that the performance feedback used in this study was implemented by university-based researchers. No studies have examined the feasibility of teachers implementing performance feedback for writing fluency in classrooms. Exploration of teachers implementing this intervention should be explicitly studied given the high likelihood of issues surrounding intervention fidelity (Kratochwill & Shernoff, 2003/2004).

Implications for practice and future directions

Given the wide variety of instructional practices that are used for writing (Cutler & Graham, 2008), a versatile evidence-based instructional intervention, like performance feedback, is a critical tool for teachers, school psychologists, and problem-solving and response to intervention teams. The important findings associated with this study indicate that performance feedback and not practice is a primary change agent and can be added to typical instructional practices. Prior to trying more intensive individualized interventions for a large number of children, it would be prudent to deliver a supplementary intervention to a whole classroom or group of students like individualized performance. Given the relative ease of implementing performance feedback interventions in the classroom with large groups of students, it is reasonable to consider implementing this type of intervention before implementing more time- and resource-intensive writing interventions. Performance feedback is not curriculum-dependent and was effective in all three schools that had varying instructional practices. Knowing that performance feedback is an important initial component to building a fluency-based intervention, future research can target school implementation issues associated with implementing fluency-based interventions as well as focus on techniques to strengthen these types of interventions. Furthermore, performance feedback could be used prior to or in conjunction with other evidence-based instructional practices (Graham et al., 2012) for written composition, like self-regulated strategy instruction. The Simple View of Writing would suggest that students receiving evidence-based instruction in planning and revising will also be more likely to have overall higher outcomes in writing if they have improved fluency in transcription skills (Berninger et al., 2006). Therefore, it may be beneficial to sequence performance feedback either prior to or in conjunction with self-regulated strategy development.

One unpredicted finding of the current study was the amount of writing fluency growth demonstrated by students assigned to the instructional control condition. A follow-up study measuring students' growth in other timed tasks during a fluency-building intervention would be helpful in determining if generalization of performance occurs. For example, a fluency-building intervention targeting writing fluency may be examined for its effect on other basic skills (e.g., basic mathematics calculation, spelling, and paragraph copying). Results from a study like that would strengthen the literature on fluency-building instruction and interventions.

The impact of performance feedback on an untargeted measure of writing quality, CWS, is encouraging. Future studies may focus on providing feedback on other writing metrics that may demonstrate greater face validity with important writing outcomes. For example, one study of students in ninth grade gave feedback on the number of adjectives used, the number of complete sentences, and the number of compound sentences (McCurdy, Skinner, Watson, & Shriver, 2008).

Although performance feedback has an impact on writing fluency outcomes, previous studies have failed to find a significant increase in general writing achievement (measured by standardized, norm-referenced tests) as a result of performance feedback alone, even when feedback was provided on a more frequent basis (Rosenthal, 2005). Additional behavioral intervention components that may enhance the effectiveness of performance feedback and subsequently increase the rate of some students' writing fluency include goal-setting and reinforcement (McCurdy et al., 2008; Morgan & Sideridis, 2006) and interdependent group-oriented contingencies (Popkin & Skinner, 2003). For example, in the area of reading, feedback combined with goal-setting or reinforcement had a greater impact than other types of interventions (i.e., repeated practice and tutoring) for improving boys' reading fluency growth (Morgan & Sideridis, 2006). In the area of mathematics, a comparison of two fluency-building interventions indicated that the type of intervention interacted with initial level of fluency in producing different rates of growth for students performing above the 15th percentile for their grade than students performing below the 15th percentile (Codding et al., 2007). Future research in the area of writing should consider conducting comparative studies to examine whether other fluency-building interventions differentially affect students' writing fluency as a function of initial level of fluency. In summary, it seems prudent for school-based practitioners to ensure that feedback in writing fluency is occurring for all students and that students who are at-risk for significant writing difficulties receive a combination of fluency-building procedures in order to at least prevent the gap in writing fluency from widening.

Acknowledgments

This study was supported, in part, by a Graduate Student Research Award from the National Association of School Psychologists in 2009 to the first author. Y. Petscher's contribution was supported, in part, by the Institute of Education Sciences' Reading for Understanding Consortium R305F100005 and Learning Disabilities Research Centers Consortium of the National Institute of Child Health and Human Development P50 HD052120.

The authors would like to acknowledge the students and teachers participating in the study and the tireless graduate and undergraduate students who assisted with the data collection and scoring.

References

Abbott RD, Berninger VW. Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. Journal of Educational Psychology. 1993; 85:478–508.

American Psychological Association Presendential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. American Psychologist. 2006; 61(4):271–285. [PubMed: 16719673]

Ardoin SP, Christ TJ, Morena LS, Cormier DC, Klingbeil DA. A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rules. Journal of School Psychology. 2013; 51:1–18. [PubMed: 23375170]

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series “B” Methodological. 1995; 57:289–300.

Berninger VW, Fuller F. Sex differences in orthographic, verbal, and compositional fluency: Implications for assessing writing disabilities in primary grade children. Journal of School Psychology. 1992; 30:363–382.

Berninger VW, Nielson KH, Abbott RD, Wijsman E, Raskind W. Sex differences in severity of writing and reading disabilities. Journal of School Psychology. 2008; 46:151–172. [PubMed: 19083355]

Berninger VW, Rutberg JE, Abbott RD, Garcia N, Anderson-Youngstrom M, Brooks A, et al. Tier 1 and Tier 2 early intervention for handwriting and composing. Journal of School Psychology. 2006; 44:3–30.

Berninger VW, Vaughan KB, Graham S, Abbott RD, Abbott SP, Rogan LW, et al. Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition. Journal of Educational Psychology. 1997; 89:652–666.

Berninger VW, Vaughan K, Abbott RD, Begay K, Coleman KB, Curtin G, et al. Teaching spelling and composition alone and together: Implications for the simple view of writing. Journal of Educational Psychology. 2002; 94(2):291.

Bramlett RK, Murphy JJ, Johnson J, Wallingsford L. Contemporary practices in school psychology: A national survey of roles and referral problems. Psychology in the Schools. 2002; 39:327–335. http://dx.doi.org/10.1002/pits.10022.

Burchinal MR, Nelson L, Poe M. Growth curve analysis: An introduction to various methods for analyzing longitudinal data. Monographs of the Society for Research in Child Development. 2006; 71:65–87.

Camarata S, Woodcock R. Sex difference in processing speed: Developmental effects in males and females. Intelligence. 2006; 34:231–252.

Christ TJ. Short-term estimates of growth using Curriculum-Based Measurement or oral reading fluency: Estimating standard error of the slope to construct confidence intervals. School Psychology Review. 2006; 35:128–133.

Christ TJ, White MJ, Ardoin SP, Eckert TL. Curriculum based measurement of reading: Consistency and validity across best, fastest, and question reading conditions. School Psychology Review. 2013; 42(4):415–436.

Christ TJ, Zopluoglu C, Monaghen BD, Van Norman ER. Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. Journal of School Psychology. 2013; 51:19–58. [PubMed: 23375171]

Codding RS, Shiyko M, Russo M, Birch S, Fanning E, Jaspen D. Comparing mathematics interventions: Does initial level of fluency predict intervention effectiveness? Journal of School Psychology. 2007; 45:603–617.

Cohen, J. Statistical power analyses for the behavioral sciences. 2. Hillsdale NY: Lawrence Erlbaum Associates; 1988.

Coker D. Impact of first-grade factors on the growth and outcomes of urban schoolchildren’s primary-grade writing. Journal of Educational Psychology. 2006; 98:471–488.

Cronbach, LJ. Research on classrooms and schools: Formulation, questions, design, and analysis. Stanford, California: Stanford Evaluation Consortium; 1976.

Cutler L, Graham S. Primary grade writing instruction: A national survey. Journal of Educational Psychology. 2008; 100:907–919.

Daly EJ, Hintze JM, Hamler KR. Improving practice by taking steps toward technological improvements in academic intervention in the new millennium. Psychology in the Schools. 2000; 37:61–72.

Daly EJ III, Martens BK, Hamler KR, Dool EJ, Eckert TL. A brief experimental for identifying instructional components needed to improve oral reading fluency. Journal of Applied Behavior Analysis. 1999; 32:83–94.

Deno SL. Developments in curriculum-based measurement. Journal of Special Education. 2003; 37:184–192.

Deno SL, Fuchs LS, Marston D, Shin J. Using curriculum-based measurement to establish growth standards for students with learning disabilities. School Psychology Review. 2001; 30:507–524.

Deno SL, Marston D, Mirkin PK. Valid measurement procedures for continuous evaluation of written expression. Exceptional Children. 1982; 48:368–371.

Deno, SL.; Mirkin, PK.; Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests. Vol. IRLDRR-22. Minneapolis, Minnesota: University of Minnesota, Institute for Research on Learning Disabilities; 1980.

Eckert, TL.; Codding, RM.; Truckenmiller, AJ.; Rheinheimer, JL. Improving children's fluency in reading, mathematics, spelling, and writing: A review of evidence-based academic interventions. In: Akin-Little, A.; Little, SG.; Bray, MA.; Kehle, TJ., editors. Handbook of behavioral interventions in the school. Washington, DC: American Psychological Association; 2009. p. 111-124.

Eckert TL, Hintze JM. Behavioral conceptions and applications of acceptability: Issues related to service delivery and research methodology. School Psychology Quarterly. 2000; 15:123–148. http://dx.doi.org/10.1037/h0088762.

Eckert, TL.; Lovett, BJ.; Rosenthal, BD.; Jiao, J.; Ricci, LJ.; Truckenmiller, AJ. Classwide instructional feedback: Improving children's academic skill development. In: Randall, S., editor. Learning disabilities: New research. Hauppauge, NY: Nova Science; 2006. p. 271-285.

Eckert, TL.; Truckenmiller, AJ.; Rheinheimer, JL.; Perry, LJ.; Koehler, JL. Improving children's academic achievement: Benefits and barriers associated with fluency-based interventions. In: Molina, DH., editor. School psychology: 21st century issues and challenges. Hauppauge, NY: Nova Science; 2008. p. 327-343.

Elliott, SN.; Witt, JC.; Kratochwill, TR. Selecting, implementing, and evaluating classroom interventions. In: Stoner, G.; Shinn, MR.; Walker, HM., editors. Interventions for achievement and behavior problems. Silver Spring, MD: National Association of School Psychologists; 1991. p. 99-135.

Espin CA, Skare S, Shin J, Deno SL, Robinson S, Benner B. Identifying indicators of written expression proficiency for middle school students. Journal of Special Education. 2000; 34:140–153.

Fearrington JY, Parker PD, Kidder-Ashley R, Gagnon SG, McCane-Bowling S, Sorrell CA. Gender differences in written expression curriculum-based measurement in third- through eighth-grade students. Psychology in the Schools. 2014; 51(1):85–96. http://dx.doi.org/10.1002/pits.21733.

Foegen A, Jiban C, Deno S. Progress monitoring measures in mathematics: A review of the literature. Journal of Special Education. 2007; 41:121–139.

Francis DJ, Santi KL, Barr C, Fletcher JM, Varisco A, Foorman BF. Form effects on the estimation of students' oral reading fluency using DIBELS. Journal of School Psychology. 2008; 46:315–342. [PubMed: 19083362]

Fuchs L. Assessing intervention responsiveness: Conceptual and technical issues. Learning Disabilities Research & Practice. 2003; 18:172–186.

Fuchs D, Fuch LS. Introduction to response to intervention: What, why and how valid is it? Reading Research Quarterly. 2006; 41:93–99.

Graham, S.; Bollinger, A.; Booth Olson, C.; D'Aoust, C.; MacArthur, C.; McCutchen, D., et al. Teaching elementary school students to be effective writers: A practice guide (NCEE 2012-4058). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; 2012. (Retrieved from ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch)

Graham S, Harris KR. It can be taught, but it does not develop naturally: Myths and realities in writing instruction. School Psychology Review. 1997; 26:414–424.

Graham S, Perin D. A meta-analysis of writing instruction for adolescent students. Journal of Educational Psychology. 2007; 99(3):445.

Graham S, Harris KR, Fink-Chorzempa B. Primary grade teachers' instructional adaptations for struggling writers: A national survey. Journal of Educational Psychology. 2003; 95:279–292.

Graham, S.; Harris, KR.; Hebert, M. Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report. Washington, DC: Alliance for Excellent Education; 2011.

Graham S, Harris KR, MacArthur C, Fink B. Primary grade teachers' theoretical orientations concerning writing instruction: Construct validation and a nationwide survey. Contemporary Educational Psychology. 2002; 27:147–166.

Gresham, FM. Evolution of the response-to-intervention concept: Empirical foundation and recent developments. In: Jimerson, SR.; Burns, MK.; Van Der Heyden, AM., editors. Handbook of response to intervention. New York: Springer; 2007. p. 10-24.

Griffiths AJ, Van Der Heyden AM, Parsons L, Burns M. Practical applications of response to intervention research. Assessment for Effective Intervention. 2006; 32:50–57.

Hammill, DD.; Larsen, SC. Test of Written Language. 3. Austin, TX: PRO-ED; 1996.

Hammill, DD.; Hresko, WD. Comprehensive Scales of Student Abilities. Austin, TX: PRO-ED; 1994.

Hayes, JR.; Flower, LS. Identifying the organization of writing processes. In: Gregg, LW.; Steinberg, ER., editors. Cognitive processes in writing. Hillsdale, NJ: Erlbaum; 1980.

Jones D, Christenson CA. Relationship between automaticity in handwriting and students' ability to generate written text. Journal of Educational Psychology. 1990; 91:44–49.

Kazdin AE. Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. American Psychologist. 2008; 63:146–159. [PubMed: 18377105]

Kim, Y-S.; Al Otaiba, S.; Wanzek, J.; Gatlin, B. Toward an understanding of dimensions, predictors, and the gender gap in written composition. Journal of Educational Psychology. 2014. http://dx.doi.org/10.1037/a0037210 (Advance online publication)

Kratochwill TR, Shernoff ES. Evidence-based practice: Promoting evidence-based interventions in school psychology. School Psychology Quarterly. 2003/2004; 18:389–408. (School, Psychology Review, 33, 34–48).

Kratochwill TR, Stoiber KC. Empirically supported interventions and school psychology: Conceptual and practical issues—Part II. School Psychology Quarterly. 2000; 15:233–253. http://dx.doi.org/10.1037/h0088786.

Kratochwill TR, Stoiber KC. Evidence-based interventions in school psychology: Conceptual foundations of the Procedural and Coding Manual of Division 16 and Society for the Study of School Psychology Task Force. School Psychology Quarterly. 2002; 17:341–389. http://dx.doi.org/10.1521/scpq.17.4.341.20872.

Malecki CK, Jewell J. Developmental, sex, and practical considerations in scoring curriculum-based measurement writing probes. Psychology in the Schools. 2003; 40:379–390.

Marston, D.; Deno, S.; Tindal, G. A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress. Vol. IRLD-RR126. Minneapolis, Minnesota: University of Minnesota, Institute for Research on Learning Disabilities; 1983. p. 29

Marston, D.; Lowry, L.; Deno, S.; Mirkin, PK. Spelling and written expression: A longitudinal study. Vol. IRLD-RR-49. Minneapolis, Minnesota: University of Minnesota, Institute for Research on Learning Disabilities; 1981.

Martens, BK.; Daly, EJ.; Begeny, JC.; Van Der Heyden, A. Behavioral approaches to education. In: Fisher, W.; Piazza, C.; Roane, H., editors. Handbook of applied behavior analysis. New York: Guilford; 2011. p. 385-401.

Martin DJ, Hoover HD. Sex differences in educational achievement: A longitudinal study. Journal of Early Adolescence. 1987; 7:65–83.

McCurdy M, Skinner C, Watson S, Shriver M. Examining the effects of a comprehensive writing program on the writing performance of middle school students with learning disabilities in written expression. School Psychology Quarterly. 2008; 23:571–586.

McMaster, KL.; Campbell, H. Advances in monitoring progress in writing. Paper presented at the progress monitoring conference; Minneapolis, MN. 2006.

McMaster KL, Campbell H. New and existing curriculum-based writing measures: Technical features within and across grades. School Psychology Review. 2008; 37:550–566.

McMaster KL, Espin C. Technical features of curriculum-based measurement in writing: A literature review. Journal of Special Education. 2007; 41:68–84.

Mirkin, PK.; Deno, SL.; Fuchs, SL.; Wesson, C.; Tidal, G.; Marston, D., et al. Procedures to develop and monitor progress on IEP goals. Minneapolis, Minnesota: University of Minnesota, Institute for Research on Learning Disabilities; 1981.

Morgan PL, Sideridis GD. Contrasting the effectiveness of fluency interventions for students with or at risk for learning disabilities: A multilevel random coefficient modeling meta-analysis. Learning Disabilities Research & Practice. 2006; 21:191–210.

National Association of School Psychologists. Prevention and intervention research in the school (position statement). Bethesda, MD: Author; 2005.

National Center for Education Statistics. The nation's report card: Writing 2011 (NCES 2012-470). Washington, D.C: Institute of Education Sciences, U.S. Department of Education; 2012.

National Commission on Writing. The neglected "R": The need for a writing revolution. The College Board; 2003.

National Mathematics Advisory Panel. Foundations for success: The final report of the National Mathematics Advisory Panel. Jessup, MD: U.S. Department of Education; 2008.

Pearson Education, Inc. AIMSweb progress monitoring and RTI system. Upper Saddle River, NJ: Author; 2009. (http://www.aimsweb.com)

Pearson Education, Inc. AIMSweb national norms technical documentation. Upper Saddle River, NJ: Author; 2012. (http://www.aimsweb.com)

Persky, HR.; Daane, MC.; Jin, Y. The nation's report card: Writing 2002 (NCES 2003-529). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; 2003.

Peugh JL. A practical guide to multilevel modeling. Journal of School Psychology. 2010; 48:85–112. http://dx.doi.org/10.1016/j.jsp.2009.09.002. [PubMed: 20006989]

Popkin J, Skinner CH. Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. School Psychology Review. 2003; 32:282–295.

Powell-Smith, KA.; Shinn, MR. Administration and scoring of written expression curriculum-based measurement for use in general outcome measurement. Eden Prairie, MN: Edformation; 2004.

Raudenbush SW. Comparing personal trajectories and drawing casual inferences from longitudinal data. Annual Review of Psychology. 2001; 52:501–525.

Rosenthal, BD. Improving elementary-aged children's written expression skills: A comparison of improvement based on performance feedback frequency. Syracuse, NY: Syracuse University; 2005. (Unpublished doctoral dissertation)

Salahu-Din, D.; Persky, H.; Miller, J. The Nation's Report Card: Writing 2007 (NCES 2008–468). National Center for Education Statistics. U.S.: Institute of Education Sciences; 2008.

SAS Institute Inc. SAS 9.1.3 Help and documentation. Cary, NC: SAS Institute; 2004.

Schochet, Peter Z. Technical methods report: Guidelines for multiple testing in impact evaluations (NCEE 2008–4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; 2008.

Shapiro, ES. Academic skill problems: Direct assessment and intervention. New York: Guilford; 2004.

Shinn, MR. Administration and scoring of mathematics computation curriculum-based measurement and math fact probes for use with AIMSweb. Eden Prairie, MN: Edformation; 2004.

Shinn, MR.; Ysseldyke, J.; Deno, S.; Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving. Vol. IRLD-RR-71. Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities; 1982.

Singer, JD.; Willett, JB. Applied longitudinal analysis: Modeling change and event occurrence. New York: Oxford University Press; 2003.

Stebbe Rowe S. Reading intervention and assessment acceptability: A literature review. School Psychology Forum. 2012; 6:77–88.

Stoiber KC, Kratochwill TR. Empirically supported interventions and school psychology: Rationale and methodological issues — Part I. School Psychology Quarterly. 2000; 15:75–105.

Truckenmiller, AJ. Utilizing group performance feedback as a class-wide writing intervention for elementary students. Syracuse, NY: Syracuse University; 2007. Unpublished master's thesis

Van Houten R. The performance feedback system: Generalization of effects across time. Child Behavior Therapy. 1979; 3:219–236.

Van Houten R, Hill S, Parsons M. An analysis of a performance feedback system: The effects of timing and feedback, public posting, and praise upon academic performance and peer interaction. Journal of Applied Behavior Analysis. 1975; 8:449–457. [PubMed: 16795508]

Van Houten R, Morrison E, Jarvis R, McDonald M. The effects of explicit timing and feedback on compositional response rate in elementary school children. Journal of Applied Behavior Analysis. 1974; 7:547–555. [PubMed: 16795479]

Wolf MM. Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. Journal of Applied Behavior Analysis. 1976; 11:203–214. http://dx.doi.org/10.1901/jaba.1978.11-203. [PubMed: 16795590]

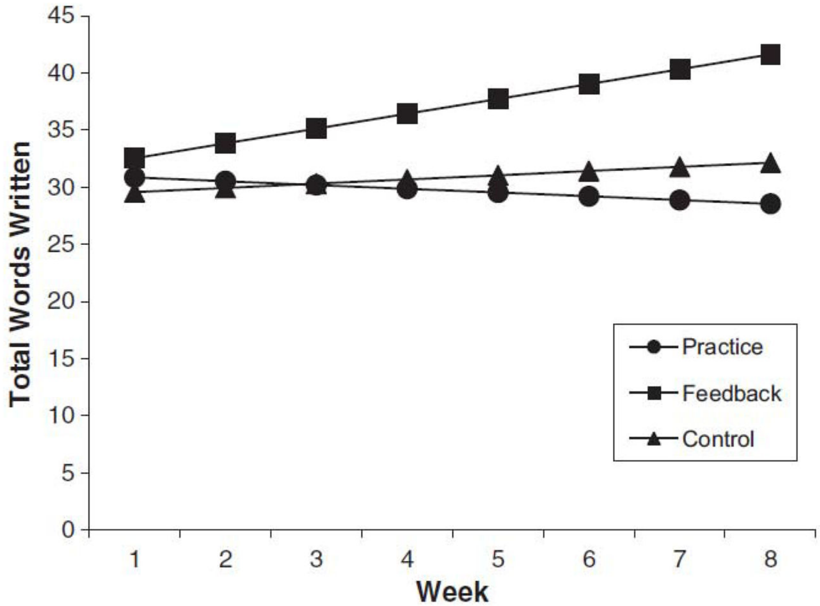

Fig. 1.
Model estimates of the growth trajectory for students' TWW by instructional practices.

Table 1

Demographic characteristics of participating students (N = 133).

Characteristic	Conditions						Total sample			
	Performance feedback		Practice-only		Instructional control					
	%	**(*n*)**	**%**	**(*n*)**	**%**	**(*n*)**	**%**	**(*n*)**	χ^2	***p***
Sex										
Girls	59	(27)	46	(18)	56	(27)	54	(72)	1.71	.43
Boys	41	(19)	54	(21)	44	(21)	46	(61)		
Ethnicity/race										
African American/Black	76	(34)	80	(32)	71	(35)	74	(99)	8.36	.21
Native American	0	(0)	5	(2)	0	(0)	2	(2)		
White	18	(8)	15	(6)	20	(10)	18	(24)		
Hispanic/Latino	7	(3)	0	(0)	8	(4)	5	(7)		
Receiving special education services	17	(8)	15	(3)	18	(9)	15	(20)	2.43	.30
Age	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*F*	*p*
	8.78	(.39)	8.77	(.37)	8.93	(.44)	8.85	(.41)	1.71	.19

Table 2

Baseline academic skills (N = 133).

	Standardized measure		Curriculum-based measurement[b]			
	Spontaneous Writing Quotient[a]		TWW		CWS	
	M	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
School						
A	96.72	(21.89)*	31.77	(13.59)	26.40	(12.77)
B	82.52	(32.42)	30.12	(11.13)	26.07	(11.16)
C	71.04	(37.14)	32.85	(12.17)	28.70	(12.73)
Condition						
Feedback	83.36	(33.28)	31.67	(12.63)	26.02	(12.75)
Practice-only	87.41	(30.35)	31.23	(12.26)	27.64	(11.27)
Control	82.70	(32.89)	31.56	(12.65)	27.04	(12.64)
Sex						
Girls	85.37	(30.96)	33.72	(12.79)*	29.18	(12.52)*
Boys	82.72	(33.94)	28.89	(11.58)	24.13	(11.35)
Total sample	84.31	(32.05)	31.50	(12.44)	26.86	(12.22)

Note. The TOWL-3 Spontaneous Writing Quotient is based on a standardized mean of 100 and standard deviation of 15.

[a] $n = 108$.

[b] $n = 133$.

[*] $p < .05$.

Table 3

Multilevel prediction models for total words written.

Estimate	Unconditional growth model		Model testing instructional effects		Model adding sex effects	
Fixed effects						
Final status						
Intercept	34.59	(1.37)**	32.15	(1.99)**	26.58	(2.46)**
Feedback			9.48	(3.09)**	12.70	(4.59)**
Practice			−3.61	(2.84)	0.38	(3.60)
Girls					10.09	(3.58)**
Feedback × Female					−6.13	(6.05)
Practice × Female					−6.73	(5.40)
Rate of change						
Time	0.51	(.15)**	0.37	(0.24)	−0.08	(0.42)
Feedback			0.93	(0.34)**	1.27	(0.63)*
Practice			−0.70	(0.33)*	−0.08	(0.49)
Feedback × Female					−0.65	(0.72)
Practice × Female					−1.21	(0.65)
Feedback vs. practice						
Final status			13.09	(3.10)**		
Rate of change			1.63	(0.32)**		
Random effects						
Intercept	196.59**		162.10**		156.12**	
Slope	0.65*		0.21		0.20	
Covariance	10.28**		6.19**		5.92*	
Residual	75.75**		71.35**		71.39**	
Goodness of fit: χ^2	5493.0		5461.3		5438.1	
AIC	5501.0		5469.3		5446.1	
AICC	5501.1		5469.4		5446.2	
BIC	5512.6		5480.9		5457.7	

Note. Standard errors of parameter estimates are denoted in parentheses.

* $p < .05$.

** $p < .01$.

Table 4

Multilevel models for correct writing sequences.

Estimate	Unconditional growth model		Model testing instructional effects		Model adding sex effects	
Fixed effects						
Final status						
Intercept	29.58	(1.21)**	28.57	(1.88)**	24.60	(2.78)**
Feedback			5.25	(2.5)*	6.39	(3.84)
Practice			−3.39	(2.67)	−1.76	(3.76)
Rate of change						
Time	0.46	(0.13)**	0.41	(0.24)	0.13	(0.36)
Feedback			0.59	(0.31)	0.76	(0.47)
Practice			−0.63	(0.32)	−0.21	(0.46)
Feedback vs. practice						
Final status			8.64	(2.55)**		
Rate of change			1.22	(0.29)**		
Random effects						
Intercept	105.28**		102.88**		98.78**	
Residual	63.30**		61.76**		61.89**	
Goodness of fit: χ^2	5385.2		5358.2		5335.7	
AIC	5389.2		5362.2		5339.7	
AICC	5389.2		5362.2		5339.7	
BIC	5395.0		5367.9		5345.5	

Note. Standard errors of parameter estimates are denoted in parentheses.

* $p < .05$.

** $p < .01$.

Table 5

Student ratings of procedural acceptability.

Procedures associated with CBM-WE (N =107)	Feedback		Practice		Control		Total	
	M	**(*SD*)**	***M***	**(*SD*)**	***M***	**(*SD*)**	***M***	**(*SD*)**
How much do you like writing stories with us each week?	4.12	(1.50)	4.48	(1.01)	3.71	(1.40)	4.07	(1.38)
How much do you like being timed while you are writing your stories with us?	2.00	(1.70)	2.89	(1.65)	1.98	(1.56)	2.22	(1.67)
Were there any times you didn't want to write a story with us?	3.59	(1.77)	4.04	(1.48)	3.00	(1.60)	3.48	(1.67)
Were there any times when you wished you could write more stories with us?	3.79	(1.79)	3.67	(1.52)	3.00	(1.73)	3.46	(1.73)
Do you think your writing has improved?	4.33	(1.30)	4.19	(1.24)	4.39	(0.92)	4.32	(1.15)
Do you think your writing has gotten worse?	4.33	(1.46)	4.37	(1.28)	4.22	(1.37)	4.30	(1.37)
Performance feedback condition (n = 39)				*M*				(*SD*)
How much did you like being told how many words you wrote?				4.26				(1.45)
How much do you think it helps when you are told how many words you wrote?				4.33				(1.30)

Note. Answers were based on a Likert-type scale with 1 = not at all, and 5 = very, very much. Items 3 and 6 were reversed scored so that higher numbers represent higher acceptability.