The Psychological Record, 2009, 59, 3–20

FLUENCY TRAINING A WRITING SKILL: EDITING FOR CONCISION

Marshall L. Dermer, Shannon L. Lopez, and Paul A. Messling III
University of Wisconsin–Milwaukee

The goal of this study was to design and evaluate fluency-based training units to help students eliminate inconcision. Participants first completed a 1.5-hr lesson on writing concisely and then a 5-min test during which they edited sentences containing inconcise text from the training units. Subsequently, participants were randomly assigned to an intervention package that included fluency training or an intervention package that included control activities. After about 5 weeks, all participants retook the test. In 2 such experiments, relative to their initial rates, fluency-trained participants improved more than did control participants; the improvement for fluency-trained participants was often maintained for 5 weeks. The findings' limitations are discussed, as well as their implications for enhancing student writing and further research.

For obvious reasons, writing is important. So, its instruction continues from the first year of grammar school through the first year of college and is supported by many resources. Among those resources, in the United States, are the dozen periodicals of The National Council of the Teachers of English (2007) and even a few articles published by behavior analysts. Behavior analysts have targeted such behaviors as spelling (Cuvo, Ashley, Marso, Zhang, & Fry, 1995); low number or rates of word or sentence production (Ballard & Glynn, 1975; Porritt, Burt, & Poling, 2006; Van Houten, Morrison, Jarvis, & McDonald, 1974); and aspects of writing, including creativity and originality (Glover, 1979) as well as quality (Ballard & Glynn, 1975; Van Houten et al., 1974). Behavior analytic treatment packages have included preparatory activities (Glover, 1979); timed and untimed practice (Cuvo et al., 1995; Glover & Gary, 1976; Van Houten et al., 1974); and, of course, consequences such as reinforcement and feedback (Ballard & Glynn, 1975; Brigham, Graubard, & Stans, 1972; Glover, 1979; Glover & Gary, 1976; Porritt et al., 2006). Despite all the resources devoted to writing instruction, by the time students enter college many write poorly, even students attending elite universities (Bartlett, 2003). Indeed, graduate student and professional writing could be much improved (Silvia, 2007).

Special thanks are due the students of our applied behavior analysis classes for participating and offering suggestions during all phases of our work, Victoria Steinbach for helping score tests, and Noah S. Dermer for writing our software.

Send correspondence to Marshall Lev Dermer, Department of Psychology, University of Wisconsin–Milwaukee, PO Box 413, Milwaukee, WI 53201. E-mail: dermer@uwm.edu

The papers of many college students, particularly upper-division students, are often written in a graceless style (see Williams, 2007). But what is style? If we can distinguish between the primary functions of some behavior or behavioral product and other functions, then *style* refers to those features of a behavior or a behavioral product that may vary while maintaining its primary functions. A prevalent, undesirable feature of undergraduate writing is inconcision. An inconcise sentence, whether short or long, is one with needless words.

Helping students write more concisely has been our research goal. In particular, we have been exploring the utility of the "parts to whole" training strategies of applied behavior analysis (see, e.g., Cooper, Heron, & Heward, 2007; Sulzer-Azaroff & Mayer, 1991). We have provided students with plenty of practice and feedback regarding editing of "parts" (inconcise short phrases, clauses, and sentences that we have written) and then examined whether they could appropriately edit "wholes" (longer and often more complex inconcise texts that either we or they have written).

Although a concise sentence is without needless words, it is difficult to define concision more precisely. There are, however, some principles that facilitate revision. For example, "Use the verb (I *respect* you) rather than its nominalization (I have *respect* for you)" and "Delete redundant words (first ~~and foremost~~)." When we could identify such principles, we had students edit multiple exemplars (e.g., Fox & Ghezzi, 2003). But not all editing can be so addressed:

> I can offer no principle that tells you when to replace a phrase with a word, much less give you the word. I can point out only that you often can, and that you should be alert for opportunities to do so—which is to say, try. (Williams, 2007, p. 115)

So, if you write "due to the fact that" and then replace the phrase with "because," your editing is not likely an instance of some generalized, higher order operant or principle (Healy, Barnes-Holmes, & Smeets, 2000; Tiemann & Markle, 1990) but is presumably due to direct training.

We found examples of inconcise writing by visiting many Web sites, consulting a dictionary of inconcision (Fiske, 1996), and scrutinizing hundreds of pages written by undergraduates. We used these examples to create frames so participants could practice recognizing and editing wordy text while preserving the text's meaning. We organized the frames into instructional units. Early units required editing inconcise phrases, clauses, and sentences; later units required reviewing one or more sentences and editing only if the texts were wordy. The multisentence frames provided practice in editing text in context.

We designed software and procedures to promote fluency, that is, high rates of accurate responding. We focused on response rates because a growing literature (e.g., Binder, 1996; Bucklin, Dickinson, & Brethower, 2001; Kubina & Wolfe, 2005) suggests that procedures that promote fluency can promote important learning outcomes, such as retention (i.e., appropriate, high-rate behavior persists long after training is completed), endurance (i.e., appropriate, high-rate behavior persists for durations greater than the durations of practice sessions), application (i.e., behavior can easily combine with other behaviors to form composites), and

adduction (i.e., behaviors combine to form new behavior with little or no additional instruction). Presumably, procedures that produce *many* fluent editing skills could help editors easily eliminate previously encountered and novel stylistic problems, long after training (Binder, 2004; see the discussion of fluency by Johnson & Layng, 1996).

During a training session, participants chose an instructional unit and session duration. Software then presented the unit's frames and recorded how participants scored their responses. When the session ended, the software presented summary statistics, including the number of correct and incorrect responses per minute. Participants graphed these rates and worked on a unit until they achieved a frequency aim, a rate achieved by a competent editor (usually the senior author). If the graphed data indicated slow or no progress, we modified the training procedure.

Much of our approach was derived from precision teaching (e.g., Lindsley, 1992; West, Young, & Spooner, 1995), but we did not follow all its procedures and conventions. For example, we neither had participants administer a 1-min test at the day's end and plot their resulting rates of corrects and incorrects on a Standard Celeration Chart nor followed all plotting conventions (Graf & Lindsley, 2002). To some extent our deviations were due to our being new to precision teaching, but they were also due to our software's tracking of a student's correct and incorrect response rates for *each* session. Like many precision teachers, however, we did collect our data in a classroom rather than laboratory, so we attempted to discern intervention effects against a background of "noise" from uncontrolled variables; but unlike most research in precision teaching (see the *Journal of Precision Teaching and Celeration*), we used a control group to rule out many plausible rival explanations for the behavior changes that appeared due to fluency training.

Besides constructing units to promote fluency, we included two introductory units that introduced participants to the software and described how to score responses for concision. We wrote these frames in a programmed learning format (e.g., Holland & Skinner, 1961).

This report describes two experiments that evaluated whether an intervention package that included fluency training helped students eliminate inconcision, relative to an intervention that controlled for many plausible rival explanations. To our knowledge, these experiments are the first to use fluency training to enhance editing. The experiments focused on whether the effects of training would generalize across contexts: When wordy text from the training units appeared in novel sentences, would students eliminate this wordiness? The experiments also explored whether the effects of training would generalize to other contexts and forms of wordiness: Would training help students eliminate wordiness from their own writing? Because we did not find that training improved students' revisions of their own writing, we focus on the first form of generalization and later discuss our attempts to affect students' own writing. Finally, we collected data only after students had completed a conventional lesson on concision so we could assess the effects of fluency training beyond those easily achieved by conventional instruction.

Method

Participants and Setting

During the spring and fall semesters of 2001, participants were, respectively, 11 and 9 psychology majors who had completed an undergraduate course in research methods in which they wrote laboratory reports. They were enrolled in an upper division course in applied behavior analysis with a 1-hr lecture on Mondays, Wednesdays, and Fridays and a 1.5-hr laboratory on Tuesdays and Thursdays.

At the semester's beginning, we informed students that the laboratory would cover two areas: instructional design and the use of behavior analysis with children with developmental disabilities. We described the work in instructional design as enhancing participants' understanding of fluency-based instruction, ability to conduct single-subject research, and writing styles. We explained that the order in which students would complete the units would be based on the requirements of our research. Although we expected all students to complete all materials, students were free to refuse or discontinue participation in the research. One student so declined, and those data are not reported.

All students first read and discussed readings that covered fluency training and precision teaching. After about 3 weeks, the research commenced.

Design and Procedure

All participants completed a pretest and about 5 days later were randomly assigned to one of two intervention packages that provided practice: editing wordy texts until fluency aims were achieved or applying behavior analysis with children with developmental disabilities. After about 5 weeks, all participants completed a posttest and then began the remaining intervention. After another 5 weeks, only the first group of fluency-trained participants completed a third, follow-up test that assessed whether gains were maintained.

Concision Lesson and Pretest

As a group, all participants first completed a 1.5-hr lesson on concision using chapters from composition texts (Bauer, 1983, Chapter 18; Warriner, 1982, Chapter 17). Participants took turns reading sections of the chapters and editing sentences aloud. After the lesson, participants completed in class, with pen and paper, the concision test—the pretest.

The test's instructions were similar across semesters, but most explicit were the fall semester's instructions. These instructions were slowly read by the senior author. Participants were to decide whether a sentence was wordy and, if it was, to make it concise by deleting, adding, or rearranging words while preserving meaning. Meaning, they were told, is preserved when any question that can be answered from the original sentence can be answered from its revision, except for questions about the original sentence's structure (see Appendix A). If participants thought a sentence was concise, they were to mark it "OK."

The concision test included 90 sentences. Each sentence did not appear in the instructional units but did contain wordy text from Unit 3 through Unit 5. The embedded texts were either taken verbatim from these units or transformed regarding tense or person. Appendix B presents the test's first 16 sentences with the wordy texts boldfaced and the answers italicized. Participants had 5 min to complete the test. For participants who had completed fluency training, these tests assessed generalization: The wordy texts that participants had edited in the instructional units could now be edited in novel contexts.

One week after administering the tests, we randomly assigned participants to one of the intervention packages. The fluency intervention was always directed by one instructor (the senior author), and the developmental disabilities intervention was directed by another instructor. For about the next 5 weeks, participants met at the same time but in different rooms with their instructors.

Fluency Training

First meeting. During the first 1.5-hr meeting, participants were introduced to the software, training units, and training and recording procedures. They were asked to work through the units in numerical order, training for about 20 min, 5 days per week, for about 5 weeks. Participants used their home computers or those in campus laboratories.

Software. A Visual Basic program presented frames.[1] Before starting the program, participants inserted a 3.5-in. disk in the computer's a:\ drive. They next activated the program, specified session duration, and selected one of nine instructional units, which activated a timer and began presenting frames. The frames for the first unit were presented in a fixed order from session to session, whereas the frames for the remaining units were presented in random order.

The participant–software interface for presenting frames is presented in the top panel of Figure 1. The top window displayed instructions and text. The middle window could display typed responses, but no participant used this window; instead, participants *thought* (covertly said) their answers. The bottom window could display answers. Besides these three windows, the interface's bottom included three buttons. The Previous button presented the previous frame, and the Stop button stopped the session and produced a screen with summary statistics. The Answer button displayed a model answer or answers and three more buttons, as depicted in the bottom panel of Figure 1. The Right button and Wrong button permitted scoring of responses as correct and incorrect. Scoring a response or removing a frame started the next frame. The Remove button removed a frame but was not used during practice sessions. Participants could reinstate removed frames by using Windows Explorer to delete a file in a:\ with the unit's name and a "skp" extension.

At a session's end, the summary screen displayed the following: the date, the unit's name, the number of frames in a unit, the number of correct responses per minute, the number of incorrect responses per minute, and the session's duration. The software also recorded these data and the session's date and time on the disk in ASCII and encrypted formats.

1 Version 1.03 of our software and other materials are available for research purposes.

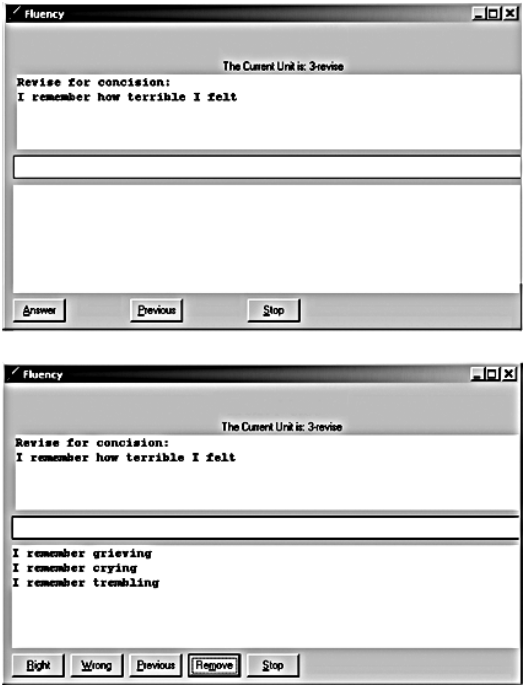

Figure 1. Participant–software interface when a new frame was presented (top panel) and after the Answer button was depressed (bottom panel.). The frame is from Unit 3.

Aims and training units. We initially specified aims based on 1-min performances that the senior author had achieved after considerable practice. We defined aims in terms of the number of net correct responses per minute, that is, the number of correct responses per minute minus the number of incorrect responses per minute. The aims were defined with respect to *silently reading* text and *thinking* responses.

We organized training material into nine units. The first and second units included the same 64 frames. Each frame, written in a programmed learning format, required single- or multiple-word responses. Because we wrote these frames in a programmed learning format, we set the aim at only 12 net correct responses per minute. For these units a higher aim may have promoted inappropriate stimulus control, such as the first few words in a frame evoking a correct response (see, e.g., Bruce, 1999, pp. 3-6 to 3-10; Fox & Ghezzi, 2003, p. 18). These units introduced participants to fluency training, described the training objectives, and taught participants how to score responses.

The third through sixth units each included about 200 frames. We set the aims for these units at 16 net correct responses per minute. Each frame presented the instruction "Revise for concision:" and inconcise text (see Figure 2). The texts, using Fiske's (1996) classification system, included, prepositional phrases (*over the duration of* → *during*), verb phrases (*is lacking in* → *lacks*), superfluous modifiers (*actually closed* → *closed*), and circumlocutions (*over the long term* → *eventually*).

The seventh through ninth units contained from 130 to 192 frames. We

initially set the aims for these units at 5 net correct responses per minute but increased the aims to 10 net correct for the fall experiment after participants began reporting such high rates and those rates were confirmed during the semiweekly tests noted below. The frames contained one or two sentences and prompted participants to edit the sentence or sentences only if participants deemed them inconcise.

For the seventh through ninth units, for every frame presenting inconcise text another frame presented nearly identical concise text. These frame "twins," consisting of a wordy example and a similar but concise nonexample, forced participants to respond to the inconcise text and not merely to the text common to the concise and inconcise versions (see Bruce, 1999, pp. 3-6 to 3-10). Moreover, preliminary work suggested that when a unit contained only one inconcise and one concise version of a frame, the previous version of a frame provided a context for responding to the frame's twin. For example, if the previous version was concise, then the next version was inconcise. To address this problem, for these units we included two inconcise and two concise versions of a frame.

Recording and graphing data. For each session, participants recorded data from the summary screen onto a summary sheet with corresponding columns. In an additional column, participants recorded the cumulative number of minutes they had spent training in a unit. As a function of these cumulative minutes (Bruce, 1999), participants graphed their rates of correct and incorrect responding on three-cycle log paper, as illustrated in Figure 2.

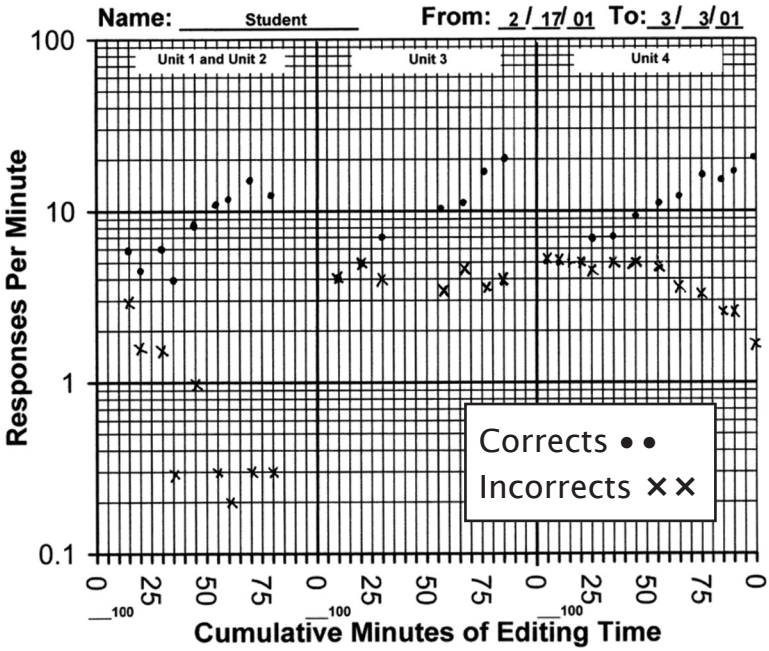

Figure 2. Graph for plotting rates of correct and incorrect responses as a function of up to 300 cumulative min of editing time. The bottommost horizontal line denotes 0 responses per minute. Corrects are indicated with dots; incorrects are indicated with x's. The data are real. Note that in the leftmost panel, the participant graphed the rates for Unit 1 and Unit 2 together because the frames for these units were identical.

Semiweekly meetings and adjusting training procedures. Each Tuesday and Thursday, for about 30 min, the instructor met with participants to verify data and review and discuss graphs. If visual inspection of a graph revealed that a participant's correct responses per minute were not increasing or incorrect responses per minute were not decreasing, then the participant either reduced the duration of sessions—say, from two 10-min sessions to four 5-min sessions—or randomly deleted 30% to 50% of the frames in a unit. When participants achieved a unit's aim with the shortened unit, they then worked with the complete unit until they achieved the unit's aim.

At the end of these meetings, the instructor individually tested participants for 1 min to check each participant's scoring. Before testing, the instructor asked participants to indicate their net rates correct for that unit. As a participant prepared to meet or beat that rate, the instructor and group often cheered. During tests, participants usually silently read texts, but they always presented their revisions out loud, which the instructor scored with the computer's mouse.

Applying Behavior Analysis with Children with Developmental Disabilities

Readings. For about the first 2 weeks, participants met with their laboratory instructor for 1.5 hrs on Tuesdays and Thursdays to read and discuss articles covering functional assessment.

Software and Procedure. For the remaining 3 weeks, participants used the 1.5-hr laboratory sessions to work with Simulations in Developmental Disabilities, a multimedia software that provides practice in applying behavior analysis with children (Desrochers, Clemmons, Grady, & Justice, 2000). Groups of two to three students worked through the program's various cases and discussed problems and strategies.

Posttest and Follow-Up Test

After about 5 weeks, the semiweekly, 1-min assessments indicated that almost all fluency-trained participants had completed Units 3 through 5, so both groups retook the concision test. Then the groups (but not the instructors) switched conditions, and about 5 weeks later only the participants who had first completed fluency training completed the concision test a third time. (For this follow-up testing, during the fall experiment participants earned 5 cents for each correctly revised sentence and lost 5 cents for each incorrectly revised sentence.)

Treatment Integrity

Although the fluency software administered the treatments, it was important to verify that the participants activated the software as per their summary sheets. So, we compared participants' summary sheets with their data in the computer-generated, encrypted files and found very high agreement. Unfortunately, though, a few participants failed to complete the requisite hours of training. Their data, however, were included to preserve random assignment.

Because the participant and not the software scored responses, we also compared the rates participants recorded on their summary sheets with those

they obtained on the semiweekly 1-min tests administered by the instructor. Although these later rates were not recorded, so correspondence cannot be quantified, the rates for these tests were comparable to those reported for training.

Scoring Concision Test and Computing Interobserver Agreement

For each sentence on the concision test, we developed model answers. For the spring experiment, we based these answers on the model answers from the training units; for the fall experiment, we also included acceptable answers from the spring experiment. Observers marked a revision correct if it preserved the original sentence's meaning and was no longer than the longest model answer.

We calculated overall agreement by randomly selecting two participants from each condition and twice scoring all their tests. For the spring and fall experiments, respectively, we so scored 37% and 43% of all the participants' tests. We calculated overall agreement using the formula [(number of agreements)/(number of agreements + disagreements)] × 100%. For the spring experiment, the average overall agreement was 85% (range: 73% to 95%), and for the fall experiment it was 94% (range: 82% to 100%).

Results

For each concision test, we subtracted participants' rates of incorrect responding from their rates of correct responding and divided by 5 min to compute their net rate of correct responding per minute. For the spring experiment, the top panel of Figure 3 presents these rates for control participants (dashed lines) and fluency-trained participants (solid lines). The slopes of all the lines from Time 1 to Time 2 are more positive for fluency-trained participants than for control participants. These differential gains are confirmed by comparing the average changes, 3.4 and .36, respectively, $t(9) = 3.74$, $p = .005$, $\eta 2 = .61$.

Consider the performances of the fluency-trained participants in greater detail. Before training, at Time 1, these participants all had higher rates of incorrects than corrects. They averaged –1.04 net correct per minute, despite having completed 1.5 hrs of instruction on editing for concision. After training, at Time 2, the reverse was true: They averaged 2.4 net correct per minute. At follow-up, Time 3, their average net rate correct declined to 1.96, but the decline was not reliable, $t(4) = .53$, $p = .62$, and exceeded their average net rate per minute at Time 1, $t(4) = 3.30$, $p = .03$, $\eta 2 = .49$.

For the fall experiment, the bottom panel of Figure 3 presents the results. Again, the slopes of all the lines from Time 1 to Time 2 are more positive for fluency-trained than for control participants. These differential gains are confirmed by comparing the average changes in net rate correct per minute, 3.48 and .2, respectively, $t(7) = 3.39$, $p = .01$, $\eta 2 = .61$.

Again, consider the performances of the fluency-trained participants. Before training, at Time 1, three of the five participants had higher rates of incorrects than corrects. The group averaged –.36 net correct per minute, despite having just completed 1.5 hrs of instruction on editing. After training, at Time 2, participants averaged 3.12 net correct per minute. At follow-up, Time 3, their average net rate correct declined to 3.00, but the decline was

not reliable, $t(4) = .42$, $p = .70$, and their average follow-up rate exceeded their average rate at Time 1, $t(4) = 6.20$, $p = .003$, $\eta^2 = .57$.

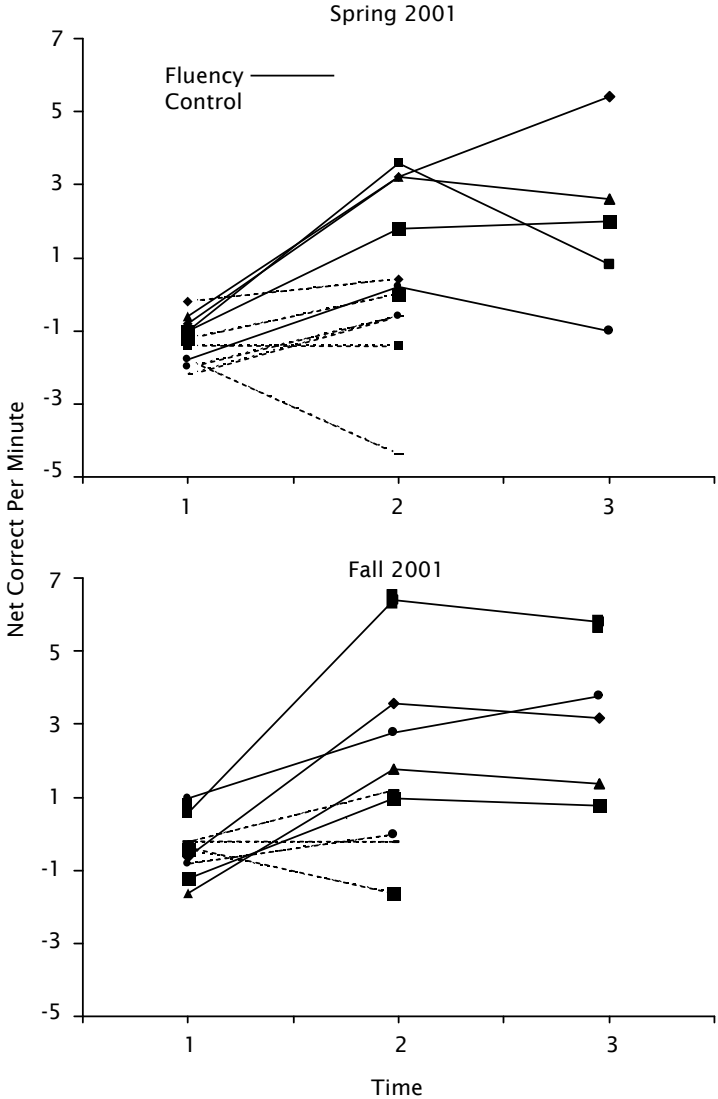

Figure 3. For the spring experiment (top panel) and the fall experiment (bottom panel): The net rate of correct responding per minute on the 5-min concision test, for fluency-trained participants (solid lines) and control participants (dashed lines) at Time 1 and Time 2. Fluency-trained participants also completed a follow-up test at Time 3. For each experiment, a unique line–symbol combination depicts each participant's performances. About 5 weeks elapsed between tests.

As noted, only during the fall experiment did we provide monetary consequences for responding on the follow-up test. Visual inspection of the changes from Time 2 to Time 3 suggests that the changes were least variable

during the fall experiment. Quantitative analysis supports this conclusion. The variance of the changes was 3.39 for the spring experiment but only .41 for the fall experiment, $F(4, 4) = 8.22$, $p = .03$.

Discussion

To evaluate our fluency-based training units, we embedded inconcise text from early units (Units 3 through 5) into novel sentences to construct a concision test. We administered this test, for 5 min, at least twice to all participants. On each administration, we asked participants to decide whether sentences were inconcise and, if so, to edit them while preserving their meaning. Before the first test, all participants had completed a 1.5-hr group lesson on editing for concision that presumably enhanced their performance when first tested. Relative to that performance, fluency-trained participants improved more than did control participants, and this improvement was evident when the test was readministered some 5 weeks after training. Moreover, this finding was replicated across two experiments, with the independent variable accounting for a quite substantial 61% of the variance in editing, from Time 1 to Time 2, for the net rate of correct responses per minute. Finally, comparisons between experiments suggest that monetary payment for responding on the follow-up test reduced variability in changes in the net rate of correct responding relative to the previous test.

These findings are encouraging but it is important to consider issues regarding internal validity, final performances, and the external validity or generality of our findings. Regarding internal validity (Campbell & Stanley, 1963), our fluency intervention differed from the control intervention in multiple ways. First, one instructor always supervised the fluency training meetings and another supervised the developmental disabilities meetings. Although we cannot be absolutely sure whether the different interventions or instructors enhanced performance, this confound is implausible. If the particular instructor alone could produce fluency, then fluency should have been immediately evident when participants *began* training with this instructor. Recall that participants trained for about 20 min daily and graphed performances, but fluency in editing did not occur immediately. Fluency was achieved gradually, as participants trained. Such a gradual progression is illustrated in Figure 2, for Unit 3 and Unit 4, for a participant from the spring experiment. Nevertheless, future research can address this confound by either having one instructor direct both conditions or counterbalancing instructors across replications of the experiment and verifying that training is effective with *each* instructor.

Other confounds include, relative to the fluency condition, the control condition involving: more reading and discussion of journal articles, more time with the laboratory instructor, more time in the laboratory, and meeting in a larger room. Like the instructor confound, these differences are implausible explanations for the greater improvement from Time 1 to Time 2 for the fluency-trained participants than for the control participants.[2] Related to the issue of internal validity is our having included fluency-training within a package of components, so we cannot specify whether it was the training alone or the training in combination with the other components that enhanced performance.

A final internal validity issue pertains to the performance of fluency-

trained participants on the follow-up test. Was that performance maintained because participants had completed fluency training about 5 weeks earlier or because they had completed the concision test twice before? We may consider the effects of completing two tests by considering the effects of completing one test, as the control participants did. Figure 3 presents such data for net rate of correct responding. Many of the lines for individual control participants show small improvements from the first to second test (from Time 1 to Time 2). On average, as noted in the Results section, the changes were not reliable. Still it is possible that a portion of the performance at Time 3 for the fluency-trained participants was due to prior testing. Indeed, the effect of repeated testing on performance is an important issue. If repeated testing produces fairly stable data before and after intervention, then single-participant designs would be feasible (Dermer & Hoch, 1999).

Also important are a number of issues relevant to external validity or generality (Campbell & Stanley, 1963). Distinct from the internal validity issue of whether an instructor alone could have enhanced editing is the external validity issue of whether the fluency training's effectiveness might depend on the one instructor who supervised such training. Although much of this training was automated, such generality can be addressed by future researchers having multiple instructors supervise fluency training.

A far more important generalization issue is whether training will enhance students' editing of *their own* writing. We explored this issue in prior, unpublished group experiments. Importantly, in these explorations, only one observer assessed writing, and statistical analyses of these data did not reveal treatment effects, so we neither proceeded to use a second observer nor assessed interobserver agreement.[3]

In these exploratory experiments, we used a design and procedures similar to those for the subsequent experiments conducted during the spring and fall of 2001, detailed in this article. Basically, participants were first to write a concise 1,000-word essay honoring someone (see, e.g., Waitz, 1995) and then revise the essay, outside of the laboratory, after the fluency training or control interventions. To assess concision, an editor examined about 200 words from the first essay and equivalent text from the revised essay. The editor, blind to experimental conditions, then revised each participant's texts for concision and compared the number of words in these revisions with the number of words in the participants' texts. Though we explored a number of measures, none appeared particularly sensitive to fluency training.

As noted in the introduction, during our spring and fall experiments of 2001 we also explored whether training enhanced participants' editing of their writing. In the fall experiment, for example, we asked participants to write, outside of the laboratory, a concise essay honoring someone. They wrote this essay after they had completed the first administration of the concision test but before we assigned participants to conditions. After participants had completed the second administration of the concision test in the laboratory,

3 We did attempt to develop a reliable method for scoring wordiness but could not achieve high levels of interobserver agreement (Banerjee, 1999). Johnson and Street (2004) also noted problems with reliably scoring responses when many variations in response are correct: "Like our counterparts in the public schools, we are finding assessment of authentic outcomes . . . to be challenging" (pp. 114–115).

we administered another test that required them to edit their own writing. The items for this test of "own writing" were wordy sentences drawn from the participant's essay. That is, each participant received an approximately 200-word test *that contained only the participant's own wordy writing*. To help participants edit, they could access dictionaries and thesauruses in the laboratory. Provided these revised sentences preserved the original sentences' meanings, we counted the number of words in the original sentences and the revised sentences. Comparing such counts appears to be quite promising, although we did not find training to reliably enhance participants' editing of their own writing.

The absence of generalization from isolated exercises to one's own writing has been a major concern of traditional English instructors and researchers. Consider these quite critical comments of approaches like ours:

> I am convinced that one reason our traditional teaching of grammar has little *transfer to writing situations* [italics added] is the underlying behaviorist learning theory. We have simply taken for granted the behaviorist ideas that practice makes perfect and that skills practiced in isolation will be learned that way and then applied as relevant. We have assumed that this is the way teaching and learning should work, despite the overwhelming evidence that it doesn't. (Weaver, 1996, p. 18; also see Connors, 2000, pp. 96, 116)

Although we would agree that the relation between good grammar skills and writing may be remote, we would not discount practice. A procedure may fail not because it is generally ineffective but because its parameters are inappropriately set. Consider, for example, how poorly an automotive engine operates when its ignition timing is incorrect.

We may have faced a similar problem in setting inadequate aims for our fluency training units. We really do not know the aims that produce such desirable outcomes as retention, endurance, application, and adduction (Johnson & Layng, 1996; Kubina & Wolfe, 2005). In this regard, consider the follow-up performances of the fluency-trained participants. They correctly edited, on average, 68% and 73% of the sentences they attempted during the spring and fall semesters, respectively. In comparison, several years after developing the units, the senior author correctly edited 95% of the sentences he attempted. Moreover, his rates were 7.2 corrects and .4 incorrects per minute. These rates were only approximated by one participant. Perhaps future research should explore the effects of higher aims in conjunction with using more of the short, wordy texts that appeared in Units 3 through 5. If participants correctly edited only about 70% of the sentences attempted on a generalization test directly related to training, could they better edit text less related to training—that is, their own writing?

Conclusions

We do not know why we failed to detect generalization beyond our test items. Was the lack of generalization due to the design of our fluency units, the need for programming additional activities that promote generalization, or just the difficulty of reliably measuring concision when there are many correct responses as in editing one's own writing?

Our goal has been to enhance editing of one's own inconcise writing.[4] Regarding this behavior, it is important to note that we have never preselected participants for high levels of inconcision. We have just assumed that most participants write inconcisely. Indeed, when we have asked participants, usually graduating seniors, to write concise essays after 1.5 hrs of concision instruction, they have used, on average, 1.8 words (range: 1.2–2.4 words) when 1 word would suffice. Although the average confirms our assumption that participants typically write inconcisely, the range suggests that researchers ought to select participants who write extremely inconcisely when assessing generalization and application (Johnson & Street, 2004).

A final methodological point concerns research design. Researchers in this area have used group analogs to single-subject designs (Ballard & Glynn, 1975; Glover & Gary, 1976; Brigham et al., 1972). Indeed, the multiple-baseline design replicated across groups (e.g., Porritt et al., 2006) could make fluency-based research feasible in university-level writing courses.

The experiments we conducted during the spring and fall of 2001 fairly clearly show that fluency training helped participants copyedit sentences with wordy text drawn from our fluency units. This effect was evident immediately after training *and* appears to have persisted on the follow-up tests administered 5 weeks later. These follow-up performances are fairly impressive. Consider, for example, the sentences that appear in Appendix B, which are quite representative of the sentences participants edited. On average, these 16 sentences include 8.9 words but their revisions include just 5.8 words.[5] So a fluency-trained participant who correctly edited 10 such sentences would use no more than 58 words (10×5.8) instead of 89 words (10×8.9). Using 58 words instead of 89 words is a 35% reduction. By the same logic, on the follow-up tests, fluency-trained participants did correctly edit, on average, about 70% of the sentences. So, in terms of editing 10 of the sentences in Appendix B, the typical participant would use no more than 67.3 words ($7 \times 5.8 + 3 \times 8.9$) instead of 89 words ($10 \times 8.9$). Using about 67 words instead of 89 words is a 24% reduction. Put another way, would you prefer reading 89 pages when 67 pages would suffice?

References

BALLARD, K. D., & GLYNN, T. (1975). Behavioral self-management in story writing with elementary school children. *Journal of Applied Behavior Analysis, 8,* 387–398.

BANERJEE, A. (1999). *On reliably scoring wordiness.* Unpublished master's thesis. University of Wisconsin–Milwaukee.

BARTLETT, T. (2003, January 3). Why Johnny can't write, even though he went to Princeton. *The Chronicle of Higher Education,* p. A39.

BAUER, M. B. (1983). *Prentice-Hall grammar and composition, Grade 12.* Englewood Cliffs, NJ: Prentice-Hall.

4 A more modest generalization issue was explored by Lopez (2004). She replicated the experimental design and interventions of our spring and fall experiments of 2001 but used a test whose wordy items formally differed from the text in the training units. Although she did not detect training effects, researchers concerned with such generalization should review Lopez's work to avoid the problems she encountered.

5 When a sentence had multiple model answers, we used the longest answer to compute the length of the average revised sentence.

BINDER, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19,* 163–197.

BINDER, C. (2004). A refocus on response-rate measurement: Comment on Doughty, Chase, and O'Shields (2004). *The Behavior Analyst, 27,* 281–286.

BRIGHAM, T. A., GRAUBARD, P. S., & STANS, A. (1972). Analysis of the effects of sequential reinforcement contingencies on aspects of composition. *Journal of Applied Behavior Analysis, 5,* 421–429.

BRUCE, G. S. (1999). *Instructional design made easy: A workbook for designing accelerated learning programs* (2nd ed.). Tucker, GA: Aubrey Daniels & Associates.

BUCKLIN, B. R., DICKINSON, A. M., & BRETHOWER, D. M. (2001). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13,* 140–163.

CAMPBELL, D. T. (1984). Can we be scientific in applied social science? *Evaluation Studies Review Annual, 9,* 26–48.

CAMPBELL, D. T., & STANLEY, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

CONNORS, R. J. (2000). The erasure of the sentence. *College Composition and Communication, 52,* 96–128.

COOPER, J. O., HERON, T. E., & HEWARD, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ : Pearson/Merrill-Prentice Hall.

CUVO, A. J., ASHLEY, K. M., MARSO, K. J., ZHANG, B. L, & FRY, T.A. (1995). Effect of response practice variables on learning spelling and sight vocabulary. *Journal of Applied Behavior Analysis, 28,* 155–173.

DERMER, M. L., & HOCH, T. A. (1999). Improving depictions of single-subject experiments in research texts written for undergraduates. *The Psychological Record, 49,* 49–66.

DESROCHERS, M. N., CLEMMONS, T., GRADY, M., & JUSTICE, B. (2000). An evaluation of Simulations in Developmental Disabilities (SIDD): Instructional software that provides practice in behavioral assessment and treatment decisions. *Journal of Technology in Human Services, 17,* 15–27.

FISKE, R. H. (1996). *Dictionary of concise writing.* Cincinnati, OH: Writer's Digest.

FOX, E. J., & GHEZZI, P. M. (2003). Effects of computer-based fluency training on concept formation. *Journal of Behavioral Education, 12,* 1–21.

GLOVER, J. A. (1979). The effectiveness of reinforcement and practice for enhancing the creative writing of elementary school children. *Journal of Applied Behavior Analysis, 12,* 487.

GLOVER, J., & GARY, A. L. (1976). Procedures to increase some aspects of creativity. *Journal of Applied Behavior Analysis, 9,* 79–84.

GRAF, S., & LINDSLEY, O. (2002). *Standard celeration charting 2002.* Youngstown, OH: Graf Implements.

HEALY, O., BARNES-HOLMES, D., & SMEETS, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior, 74,* 207–227.

HOLLAND, J. G., & SKINNER, B. F. (1961). *The analysis of behavior.* New York: McGraw-Hill.

JOHNSON, K. R., & LAYNG, T. V. J. (1996). On terms and procedures: Fluency. *The Behavior Analyst, 19,* 281–288.

JOHNSON, K. J., & STREET, E. M. (2004). *The Morningside model of generative instruction: What it means to leave no child behind.* Concord, MA: Cambridge Center for Behavioral Studies.

KUBINA, R. M., & WOLFE, P. (2005). Potential applications of behavioral fluency for students with autism. *Exceptionality, 13,* 35–44.

LINDSLEY, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25,* 51–57.

LOPEZ, S. L. (2004). *Can persons trained to fluently edit inconcise text, fluently edit novel inconcise text?* Unpublished master's thesis, University of Wisconsin–Milwaukee, Milwaukee, WI.

NATIONAL COUNCIL OF TEACHERS OF ENGLISH. (2007). Retrieved July 19, 2007, from http://www.ncte.org/pubs/journals.

PORRITT, M., BURT, A., & POLING, A. (2006). Increasing fiction writers' productivity through an Internet-based intervention. *Journal of Applied Behavior Analysis, 39,* 393–397.

SILVIA, P. J. (2007). *How to write a lot: A practical guide to productive academic writing.* Washington, DC: American Psychological Association.

SULZER-AZAROFF, B., & MAYER, G. R. (1991). *Behavior analysis for lasting change.* Fort Worth, TX: Harcourt Brace.

TIEMANN, P. W., & MARKLE, S. M. (1990). *Analyzing instructional content: A guide to instruction and evaluation* (4th ed.). Champaign, IL: Stipes.

VAN HOUTEN, R., MORRISON, E., JARVIS, R., & MCDONALD, M. (1974). The effects of explicit timing and feedback on compositional response rate in elementary school children. *Journal of Applied Behavior Analysis, 7,* 547–555.

WAITZ, G. (1995, November). Unforgettable Fred Lebow. *Reader's Digest, 147,* 116–121.

WARRINER, J. E. (1982). *English grammar and composition: Complete course* (Franklin ed.). Orlando, FL: Harcourt Brace Jovanovich.

WEAVER, C. (1996). Teaching grammar in the context of writing. *English Journal, 85,* 15–24.

WEST, R. P., YOUNG, R. K., & SPOONER, F. (1995). Precision teaching: An introduction. *Journal of Precision Teaching, 12,* 2–8.

WILLIAMS, J. M. (2007). *Style: Ten lessons in clarity and grace* (9th ed.). New York: Pearson/Longman.

Appendix A

Instructions for Concision Test

"Concision" derives from the Latin *concisus* which means to "cut up" or "strike." Striking EVERY needless word while retaining a text's meaning creates concision. For example, consider:

(a) In order to open the door, Jack retrieved his key.

and this revision:

(b) To open the door, Jack retrieved his key.

The revision preserves the original sentence's meaning because (just about) any question that can be answered from the original sentence can be answered from its revision. This is not true, however, for (c) and its revision:

(c) Due to the fact that Jack was sick he did not attend the party.

(d) Because Jack was sick he was absent.

Although (d) is shorter than (c), (d) does not preserve (c)'s meaning. In particular, the question "What did Jack miss?" (the party) cannot be answered from (d).

On this test, you are to enhance a text's concision while preserving its meaning. Editing can change a sentence's meaning, but if the meaning change is insubstantial then the editing is appropriate for this test.

You will have five minutes to edit sentences. For each sentence, if it is wordy, neatly make it concise by crossing out, adding, or rearranging words but not changing its meaning. If you can eliminate just one word without changing a sentence's meaning, do it! If you believe a sentence is concise, just mark it "OK."

Wait for the signal to turn the page, and do not skip sentences!

Appendix B

First 16 Items from Fall Semester's Concision Test with Wordy Text from Units 3 through 5 Boldfaced and Answers Italicized

1. She **was advanced in years** and walked slowly. *She was old (older) and walked slowly.*
2. **I am dubious** of what he said. *I doubt what she said.*
3. Because Frank was irresponsible, he **is no longer with us**. *Because Frank was irresponsible, he is dead (gone; was fired, terminated, dismissed).*
4. **I forgot to remember** to call Fran. *I forget to call Fran.*
5. **In the near future,** we will conquer heart disease. *Soon, we will conquer heart disease.*
6. When the bell rang, I **quickly put on my clothes**. *When the bell rang, I quickly dressed.*
7. **Over the long term,** the wicked are punished. *Eventually, the wicked are punished.*
8. I **do not**, any longer, **remember** him. *I forgot him.*
9. **To a certain extent,** I caused the accident. *I partially (sort of) caused the accident.*
10. **The decision of the** president was to hire sixty workers. *The president decided to hire sixty workers. The president hired sixty workers.*
11. **I approach** computing my Federal taxes **with fear and trembling.** *I dread computing my Federal taxes.*
12. **A couple of times** I felt elated. *Twice I felt elated. (Don't accept "sometimes.")*
13. When Sheila speaks with a Brooklyn accent, she **is putting on an act.** *When Sheila speaks with a Brooklyn accent, she is acting. Sheila's Brooklyn accent is phony. Sheila speaks with a phony Brooklyn accent.*
14. Jack and Jill **are able to** dance the cha-cha. *Jack and Jill can dance the cha-cha. Jack and Jill can cha-cha.*
15. **It was most often the case** that Sam would wash the dishes. *Usually, Sam would wash the dishes.*
16. **It is with sadness that we** must leave. *Sadly, we must leave.*