

THE BILINGUAL RESEARCH JOURNAL
Spring 1996, Vol. 20, No. 2, pp. 209-233

**PROCESS VERSUS PRODUCT WRITING
WITH LIMITED ENGLISH PROFICIENT STUDENTS**

Richard Gomez, Jr.
Texas Tech University

Richard Parker & Rafael Lara-Alecio
Texas A&M University, College Station

Leo Gomez
University of Texas, Pan American

Abstract

This study examined the effectiveness of free writing versus structured writing instruction with a group of 48 low-achieving limited English proficient (LEP) Hispanic students in an intensive 6-week summer program. Stratified random sampling ensured equivalent language skill levels in eight classrooms. Three classrooms were exposed to free writing, and five to structured writing. Weekly standardized writing samples were collected, and received nine scores of five types: (a) countable micro-indicators, (b) analytic ratings, (c) holistic scores, and (d) a productivity index. Structured Writing samples showed significant growth in five of nine scores, and Free Writing only one. Tests between treatments showed significant differences on just one score, in favor of Structured Writing. The authors emphasize the need for more empirical research.

Introduction

Instruction in English as a Second Language (ESL) has often focused on improving students' skills and abilities in speaking, listening, and reading in the target language (L2) while ignoring the development of the students' writing skills (Edelsky, 1982; Edelsky & Smith, 1989). Harris (1985) concluded that only 2% of ESL instruction was concerned with writing activities. Of this two percent, 72% was related to the mechanical aspects of writing such as syntax, punctuation, and spelling. Yet for L2 learners, writing is an essential language reinforcing skill (Hughey et al., 1983).

Corresponding to this lack of attention to writing instruction has been a neglect of research in writing compared to other skill areas (Graves, 1984). Only recently has writing for second language learners received research attention (Hillocks, 1986; Raimes, 1984). The increase in writing research has been due partly to the "back to basics" movement (Omaggio, 1986) and partly as a response to the current trend for increased state mandated testing (Graves, 1987). Yet the need for sound writing research is greater than ever. As teachers of ESL students become convinced of the importance of writing in language development, they must choose among very different instructional methods, without guidance from a sound research base. Two major instructional options which teachers must choose between are "free" writing and "structured" or "controlled" composition.

Free Writing

Free writing advocates promote the *process* of writing as being much more important than the product of writing. Writing is a tool for learning and self-discovery, not just a means to demonstrate learning (Emig, 1977; Meyers, 1983; Raimes, 1986). Proponents of free writing charge that structured writing strategies blunt this process (Edelsky & Smith, 1989).

Free writing can take the form of creative writing, diaries, journal writing, dialogue journals, writing a new version of a story, or rewriting the ending of a book (DeAvila, Duncan, & Navarrette, 1987). In free writing, topics are not prescribed and the writing itself is not graded (Hillocks, 1986). Graves (1981) and Staton (1982) both argue that

presenting topics deprives students of the right to generate and develop their own ideas. Murray (1978) asserts that writers most often do not know what they are going to write on, or even possibly what they have written. Proponents claim that free writing provides the incentive and practice necessary to write in effective and interesting ways, as well as practice in choosing topics (Edelsky, 1986; Emig, 1971; Graves, 1981; Murray, 1978 and others). They claim that structured writing reinforces counter-productive, mechanistic models of writing, causing students, especially weak students, from making improvement (Hartwell, 1984).

A free writing alternative to error feedback consists of teachers' personal written responses. Two studies on teachers' personal responses without corrective feedback (Glynn, Jerram & Tuck, 1986; Scriven & Glynn, 1983) concluded students wrote more in response. Furthermore, the students' subsequent papers were rated as more interesting and imaginative.

The most prevalent form of free writing in ESL classrooms is journal writing. Leaders in ESL pedagogy contend that journals are "...a valuable component in developing writing and reading competence in both first and second language classes" (Staton, 1982, p. 101). The benefits of journal writing and particularly on-going "dialogue journals" (between student and teacher) are widely extolled (DeAvila, Duncan, & Navarrette, 1987; Gutstein, 1983; Young, 1990; Wallace, 1987). Kreeft (1983) reported that through journal writing and daily feedback, ESL teachers were able to "customize" the writing process for each of their students' unique needs. Staton (1982) found multiple benefits in his analysis of twenty-six student/teacher dialogue journals in a sixth grade ESL class. Although journal writing is uniformly extolled, outcome evidence on student skill growth from this activity is lacking.

Structured Writing

"Structured writing" traditionally referred to drill-and practice, involving copying sentences, paragraphs, or essays, and correcting erroneous sentences (without actually writing) (Hammond, 1983). Unfortunately, classroom instruction for ESL students may follow this approach. Lessons are drill-and-practice based on hierarchical skill sequences, with few extended writing opportunities (Applebee, 1984; Hudelson, 1984; Zamel, 1987). This approach has been noted in

extended classroom observations (Applebee, 1984), in examination of ESL textbooks (Raimes, 1986), and in a review of teachers' scoring of ESL compositions (Mullen, 1980).

However, in non-ESL classrooms, a more contemporary form of structured writing has emerged. The focus on actual writing has replaced drill and practice of sub-skills. Structured writing in its more common, contemporary form is instruction in which: (a) the topic and purpose of the writing assignment is controlled or assigned; (b) student writing is judged for syntactic and lexical accuracy as well as "ideational content"; (c) students receive prompt error feedback and corrections on a limited number of targeted skills; and (d) students may be asked to make corrections in these prioritized skills. Composition research shows that exhaustive error correction is much less effective than selective error correction (Yates, 1983). Yet Hammond (1983) contends that "Students need to know how often they make mistakes so they can see the urgency of doing something about them" (p 189).

Structured writing typically receives micro-level error feedback; individual errors in usage, mechanics, sentence structure, and/or paragraph structure are marked. Overall scores may be a summation of the error counts (e.g. 85% correctly spelled words) or based on a separate analytic ratings, e.g. a 1-5 rating for "mechanics" or "idea development." The analytic score gives a sense of overall strength and weaknesses in each of a few (typically 2 to 6) main areas, e.g. mechanics and usage, vocabulary, idea development, paragraph cohesiveness, addressing the audience.

A potential advantage in global analytic ratings is that the student is less likely to become lost in individual error corrections. A second advantage is that the student can note improvement in these scores from one writing sample to the next. Finally, the teacher and student can both gauge a paper's relative strengths and weaknesses (Spandel & Stiggins, 1990).

Research on Free and Structured Writing

Hillocks' (1986) meta-analysis indicated that free writing has demonstrated only a minimal effect on the quality of writing. However, this research was conducted mainly with adults, many within college environments. Krashen (1977) contends that supportive results with

adult learners cannot be validly applied to younger learners. Krashen stated that error correction which may help adults become aware of exact rules and contexts for proper usage will not similarly improve children's written grammar. Hendrickson (1980) similarly claims that the attitude and confidence level is of critical importance in young learners, and should determine whether and to what extent their writing is corrected. Students with low confidence need more credit for the content of their ideas than for the form.

In the only meta-analysis on free and structured writing, only two often studies on free writing instruction involved LEP writers (Hillocks, 1986). One of these two studies, involving grades one through twelve, yielded effect sizes of .19 for free writers versus .15 for structured writers - a non-significant difference (Hillocks, 1986). The second of the LEP studies involved only sixth grade students. This study (Wienke, 1981) resulted in the free writing group posting a .36 effect size versus .15 for structured writing ($p < .01$). Therefore, for younger ESL students there presently appears to be little direct research support for structured writing.

These two approaches, free writing and structured writing, are common in classroom practice, yet neither is well-supported empirically for LEP students. Writing instruction often seen in ESL classrooms is an earlier version of structured writing, based on skill sequences and drill- and-practice. Contemporary structured writing, common in general education, involves extensive connected writing, and focuses on identifying and reducing errors. In contemporary structured writing teachers focus on a limited number of error types, and may score writing through micro-indicators or through general analytic ratings.

Purpose of the Study

The purpose of this study was to ascertain which teaching strategy, structured composition or free writing, is more effective in teaching writing skills to 6th grade LEP students in a short-term, intensive program. In addition, we initially investigated two levels of free writing-every day and every other day. To balance the treatment bias in measuring writing improvement, we selected three types of indices: (a) countable indicators, (b) analytic ratings, and (c) overall holistic ratings.

Method

Setting

The Office of Bilingual Education and Minority Languages Affairs (OBEMLA) funded the study which was conducted in a southeast Texas school district, within an intensive summer mathematics and language program. One hundred seven at-risk grade 5 LEP students underwent six weeks of daily (4.5 hrs. per day) instruction. A fluent bilingual teacher and instructional assistant (IA) were assigned to each of eight classrooms of approximately 12 students. Students were bused to a single school that housed the summer program.

Of the school district's 36% Hispanic students, 62% (6,000 students) are identified as LEP and enrolled in bilingual or ESL classes. Our program targeted LEP students transitioning from the fifth to sixth grade with low (below the 25th percentile on state norms) academic performance.

Participants

Once admitted, the 107 students were classified as Level I (14%), II (58%), or III (28%) in English language proficiency, based on cumulative files, standardized achievement from the past year, an individual interview, and informal assessment of their ability to understand and carry out classroom instructions. Students were then randomly assigned to classrooms, stratified by language proficiency level. Each class was therefore heterogeneous, with equivalent proportions of low, middle, and high English proficient students. Next, three classes were randomly assigned the structured writing (SW) treatment, and five the free writing (FW) treatment. Three of the FW classrooms practiced daily, and two practiced only half-time (three days per week). There were virtually no significant differences between the full-time and half-time FW groups on any analysis, so data from these five classrooms were combined for all analyses.

This study included a stratified random sample of 72 of the program's 107 students. From within each of the eight classes, nine students were randomly selected for this study, representing equal numbers of Level I, II, and III English proficiency. The final composition of treatment groups was: 27 SW, 45 FW.

Reduction of Sample Size

As the program was funded by a transitional bilingual grant, LEP students were permitted to write in either English or Spanish, depending upon their comfort level. However, unpredictable switching from one idiom to the other during the program resulted in non-comparable writing samples. We therefore eliminated twelve "code-switching" students from all analyses, eight from the FW group, and four from the SW group. This reduced the sample from 72 to 60. Friday absenteeism further reduced our sample. We maintained in the sample only the 48 students who were present for 6/6 or 5/6 of the Friday test probes. The group of students with one absence were compared separately with the full data group on significant differences of slope; none existed, justifying their inclusion in the study. The final sample was 48: 23 FW and 25 SW.

Instrumentation

We wished to measure writing growth through formats not biased toward a particular treatment. As a compromise, we selected four types of indices, of which two seemed biased toward each treatment. The four types of indices of writing quality were: (a) countable micro-indicators of quality (Percent of Correctly Spelled Words, Percent of Correct Word Sequences), (b) analytic ratings (Topic Development, Internal Organization, Conveying Meaning, Sentence Construction, Mechanics), (c) overall holistic ratings of communicative effectiveness, and (d) writing productivity (total words written). The countable micro-indicators appeared to be biased toward the SW treatment, and the holistic ratings and writing productivity index seemed biased toward the FW treatment. Analytic ratings did not seem clearly biased toward either treatment. A total of nine different scores were obtained from these four types of indices, summarized in Table 1.

Table 1
Three Types of Indices of Writing Quality, and Productivity

<i>Type Of Index</i>	<i>Indicators:</i>
Micro Indicators	% Correctly Spelled Words (%CSWd) % Correct Word Sequences (%CWSeq)
Analytic Ratings	Topic Development (Topic); Mechanics; Organization Of Thoughts (Organization); Conveying Meaning (Meaning); Sentence Constructions (Sentence).
Holistic Scoring	Overall Quality and Clarity of Communication to Reader (Holistic).
Productivity	Simple frequency count of words written (Total Words)

Micro-indicators. The micro-indicators and productivity index are curriculum based measurement (CBM) measures developed by Dr. Stan Deno and associates at the University of Minnesota (Deno, 1985). The % Correct Word Sequences (%CWSeq) measure was developed further by Dr. Gerald Tindal and associates at the University of Oregon (Tindal & Parker, 1989). %CWSeq refers to the proportion of adjacent word-pairs that are spelled correctly, and as pairs are grammatically correct and make sense within the story context. %CWSeq offers an advantage over sentence-level and phrase or clause-level indices (such as Hunt's T-Units) by permitting some credit (at the word-pair level) for even very low-skilled writers.

Analytic ratings. Analytic ratings were based on a modification of an ESL Writing Profile developed for use with bilingual students by Dr. Viola Florez at Texas A&M University (Florez & Hadaway, 1987). The rating scale and descriptors are presented in Appendix I.

Holistic Scoring. Holistic scoring of writing has multiple variations; here we followed guidelines developed and validated at University of Oregon (Parker, Tindal, & Hasbrouck, 1991). Raters were guided by a definition of good writing based on "communication effectiveness." Representative "range finder" writing samples were used to anchor the 1-5 point rating scale. The definition, scale, and training procedures have been used in previous studies by the second author, resulting in consistently strong interrater reliability and reasonable criterion-related validity (Parker, Tindal, & Hasbrouck, 1991).

Productivity. The Productivity index was simply the total number of words (correct and incorrect) written (Total Words) within the allotted time. This index also was developed and researched by Dr. Deno and associates (Deno, Marston, & Mirkin, 1982).

Inter-rater reliability. Periodic assessment over short time intervals requires good measurement reliability. One potential source of unreliability is inter-rater disagreement. Acceptable inter-rater agreement also is requisite to adequate power in statistical analyses on treatment effects.

All ratings were completed by two graduate students in bilingual education, both fluent in Spanish and English, and both experienced teachers. Following a week of training, reliability was assessed on a randomly selected set of 15 student papers from the first week's Friday testing. Because of the need for repeated measurement in the study, and the danger of "rater drift," reliability was assessed twice more during the six weeks. The first reliability results are presented in the following table.

Table 2
Interrater Reliability of Nine Indices of Writing Quality

Writing Scores	Percent Agreement	Cohen's Kappa
<i>Micro-Indicators:</i>		
% Correctly Spelled Words (%CSWd)	100%	1.0
% Correct Word Sequences (%CWSeq)	100%	1.0
<i>Analytic Scores</i>		
Topic of Composition (Topic)	90%	.86
Organization of Thoughts (Organization)	90%	.86
Conveying Meaning (Meaning)	97%	.95
Sentence Construction (Sentence)	94%	.91
Mechanics (Mechanics)	97%	.95
<i>Holistic Score</i> (Holistic)	93%	.91
<i>Productivity</i> (Total Words)	100%	1.0

Cohen's Kappa (Fleiss, 1981), interpreted as a Pearson r , is a conservative statistic for categorical data, which indicates interrater agreement beyond chance. Kappa corrects for both random chance (e.g. 20% on a 5-item scale) and chance due to a skewed sample. Skewed samples exist when a group of papers to be rated are not evenly

distributed in quality, but mainly high or low or medium. The tabled results show strong inter-rater Kappa scores for all variables, even when controlling for chance agreement (Huck, Cormier & Bounds, 1974).

Design

The design used was nonparametric analysis of trend in a short (6 observations) series. The design was replicated over 48 subjects, within two randomly created treatment groups. Predictor variables were Time and Group (FW versus SW). Dependent variables were nine writing proficiency indices within four categories: (a) "micro" or "countable" indicators, (b) analytic ratings, (c) holistic ratings, (d) writing productivity.

Treatment Groups

Free Writing (FW). Students in the FW treatment group engaged in writing activities every day, Monday through Friday. Students selected their own topics, except for standardized test probes on Fridays, and could write for as long as they wanted. Students' writing was not subjected to error corrections; rather, teachers responded to each student's writing through written comments. Students were then invited to respond to these comments in writing, thus creating a written dialogue. Students in FW classes also were encouraged to help one another, share drafts with their cohorts, and plan their essays in small groups as a social process.

For each student a series of six standardized Friday writing samples were evaluated to create the dependent measures of this study. To permit fair comparisons among classrooms and groups, Friday's topic was the same for all classes and treatments. However, free writing students were permitted some latitude on Friday's test sample; they could write for as long as they wanted, and could assist each other. Any student was permitted to write in Spanish or English.

Two FW classrooms practiced writing only half time, on Monday and Wednesday, while the other three FW classrooms practiced writing four of the five weekdays. No comparisons between full-time and half-time FW groups show significant group differences. Therefore, full and half-time treatment levels were collapsed together for this manuscript.

Structured Writing (SW). The SW classes engaged in daily structured writing, during the same time period as the FW group. In SW classes, topics were assigned by the teacher, and students wrote intensively, in nine minutes of concentrated writing time. Students were instructed to work alone and quietly during this concentrated writing time. Writing samples were subjected to error corrections by the teacher, who focused on those errors deemed most important. Students were directed to focus on avoiding those errors on their next writing sample. No dialogue was established between the teacher and students, and writing as a social process was not emphasized or encouraged. Instead, guidelines for the SW treatment reflected writing as an individual skill growth process. As in the FW group, students were permitted to write in Spanish or English.

Training and Supervision of Teachers and Assistants

Teachers and aides responsible for the FW and SW treatment groups were separately trained in advance of the study, in two, three-hour sessions, during a preservice week (also grant-funded). The FW and SW teachers and aides also received on-going direction and support from the first author throughout the duration of the study.

The first author of this study and an assistant conducted all training and supervision of both treatments. The principal author is fluently bilingual, and has eight years of experience in public schools as Bilingual and ESL teacher and department head. At the university level, he also trained and supervised Bilingual and ESL teachers for four years. Although he philosophically leans toward a free writing approach, in this study he made every effort to remain fair to both treatments. The assistant also was a university graduate student, fluently bilingual, and with several years experience in bilingual teaching.

Fidelity of Implementation

The first author and his assistant conducted daily monitoring during writing instruction to ensure fidelity of implementation. Checklists were developed to assist in monitoring and ensuring the presence of essential treatment components. Following a weekly monitoring visit, each classroom received a treatment fidelity percentage score, based on the

observational checklist. Very good fidelity of implementation was observed, with only a few minor corrective consultations required.

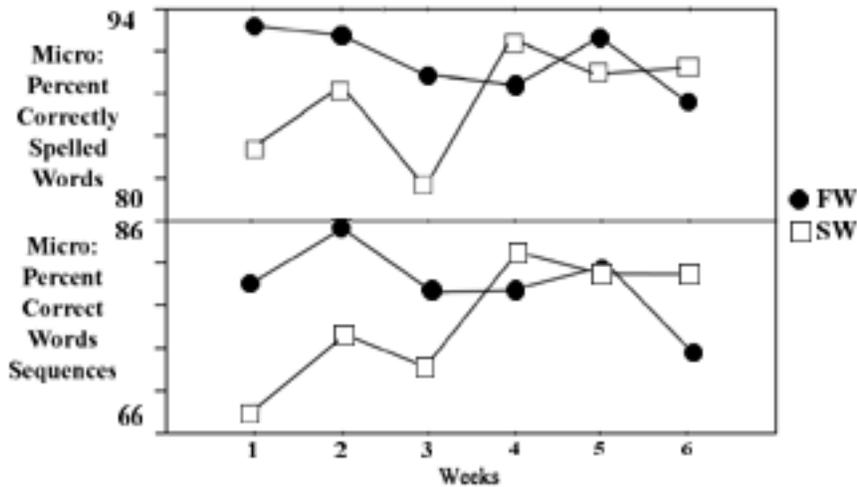
Analyses

The nonparametric Kendall's 'S' (Ferguson, 1965; Hays, 1980; Siegel, 1956) was the main analytic tool. 'S' and its better-known derivation, Kendall's Tau, measure monotonic trend in a time series. Kendall's 'S' and Tau are well-suited to single-subject time-series data with few datapoints, and permit use of ordinal and non-normal data. Tau runs from zero to one and can be interpreted loosely as "the proportion of measurements which increase over time." Tau values for any dataset will be similar to Pearson 'r's, and very similar to Spearman: Rho coefficients. Unlike Spearman's Rho, Tau reflects number of consistencies and inconsistencies in serial order, without differential weighting (Hays, 1981). The sampling distribution for 'S' is well established, and can answer the question of whether the trend in scores over time is pronounced enough to be significantly different from zero. Standard errors of measurement were derived for 'S' values, and their grouped differences between FW and SW students were statistically tested. Our analyses followed procedures in Ferguson (1965) and Darlington & Carlson (1987).

Results

The purpose of this study was to ascertain which teaching strategy, structured composition or free writing, produces greater gains in writing skills by 48 LEP students during a six week instructional program. To answer this question, we first tested whether each group's progress was significantly greater than zero. Next, we compared FW and SW group growth. Grouped data are displayed as line graphs of median scores. *Micro-Indicators.* Line graphs of median scores were plotted for FW (n=23) and SW (n=25) students. Graphs for the three micro-indicators are presented below. Note that both indices are plotted on the same "percent" scale.

Figure 1
Median Micro-indicator Scores for Free Writing (FW) and Structured Writing (SW) Groups Over Six Weeks of Instruction



Three findings are noteworthy from the Figure 1 graphs. First, the data show considerable variability from one week to the next, especially among the SW students. Second, a general improvement trend in SW students is apparent. Third, for FW students the trend appears to be in the opposite direction-deterioration. These results are largely confirmed by analyses reported in Table 3.

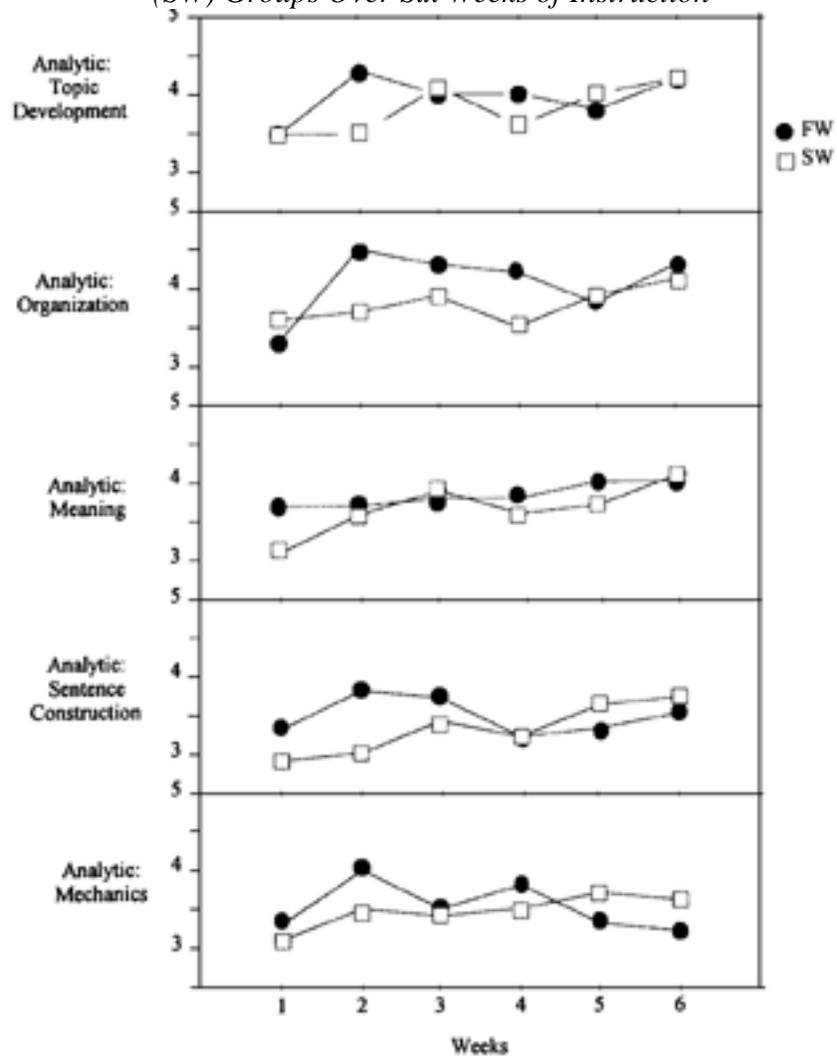
Table 3
Trend Tests and Group Differences Between FW and SW Students for Micro-Indicator Writing Scores on a Series of Six Writing Samples

Score	Group	Avg. Gain	Tau	t-test for diffs
%CSWd	FW	-6.5	-.55 (ns)	2.3 (ns)
	SW	+5%	.47 (ns)	
%CWSeq	FW	-8.7	-.33 (ns)	2.8 (.05)
	SW	+11.3	.47 (ns)	

*ns = significant

In Table 3, "Avg. Gain," expressed in raw score terms, is based on the average rate of improvement over six weeks, calculated from a "split-middle" hand-fit method (Barlow & Hersen, 1984).

Figure 2
Median Analytic Ratings for Free Writing (FW) and Structured Writing (SW) Groups Over Six Weeks of Instruction



The tabled Tau scores indicate weak-to-moderate amounts of trend in the series of scores. Negative Tau values for the FW groups confirm the negative growth observed in the graphs. However, no trends were pronounced enough for statistical significance. T-tests for FW versus SW differences between Tau scores produced no effects for %CSWd, but did yield significance ($p < .05$) for %CWSeq.

Analytic Rating Scores. Line plots also were produced for the five analytic ratings: Topic Development (Topic), Organization, Meaning, Sentence Construction (Sentence), and Mechanics. Median scores are plotted in Figure 2, all on the same scale.

As was the case with the Micro-indicators, it appears that FW scores are somewhat more variable for most analytic ratings, though very stable for "Meaning." For the SW group, a general improvement trend appears present in all scores. For the FW group, improvement is less apparent. Tau analyses are presented in Table 4.

Table 4
Trend Tests and Group Differences Between FW and SW Students for Analytic Ratings on a Series of Six Writing Samples

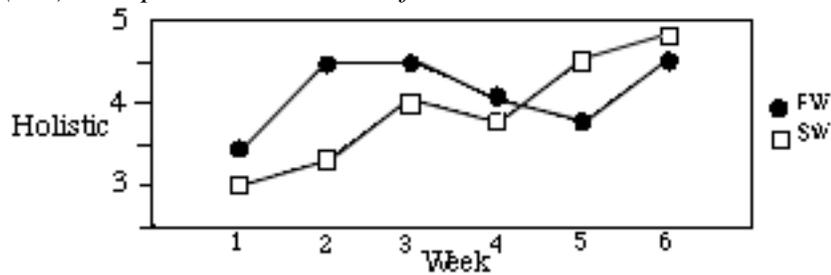
<i>Score</i>	<i>Group</i>	<i>Ave. Gain</i>	<i>Taut</i>	<i>t-test for diffs</i>
Topic	FW	.30	.14 (ns)	1.0 (ns)
	SW	.65	.69 (.05)	
Organ.	FW	.40	.00 (ns)	1.6 (ns)
	SW	.40	.55 (ns)	
Meaning	FW	.40	.89 (.01)	.19 (ns)
	SW	1.15	.69 (.05)	
Sentence	FW	-.20	-.14 (ns)	1.2 (ns)
	SW	.80	.87 (.01)	
Mechanics	FW	-.30	-.41 (ns)	1.8 (ns)
	SW	.47	.69 (.05)	

*ns = not significant

Table 4 confirms and provides details to our graph examination. Average gain was positive except for FW students on Sentence and Mechanics. Growth by SW students was large and stable enough to be statistically significant on all scores but Organization, while the FW group showed significant growth only on Meaning. However, no differences between groups were statistically significant.

Holistic Rating. Trend analysis also was performed on the Holistic rating (1-5 scale) of overall writing quality. Results are displayed graphically in Figure 3.

Figure 3
Median Holistic Ratings for Free Writing (FW) and Structured Writing (SW) Groups Over Six Weeks of instruction.



As with the Analytic Ratings, on Holistic scores the FW group displays more variability over time, and less overall progress than the SW group. These results are also presented in tabular form in Table 5.

Table 5
Trend Tests and Group Differences Between FW and SW Students for Holistic Ratings on a Series of Six Writing Samples

Score	Group	Avg. Gain	Tau	t-test for diffs
Holistic	FW	.3	.15 (ns)	.83 (ns)
	SW	1.2	.87 (.01)	

*ns = not significant

Table 5 indicates the positive SW group improvement in Holistic scores was significant beyond chance levels ($p < .01$), whereas the improvement of the FW group was not. The SW group gains were large, averaging 1.2 points on a five-point scale. However, the t-test for Tau differences was not significant.

Productivity Rating. Finally, we tested a simple index of productivity for trend, the total number of words written (Total Words). These results are displayed graphically in Figure 4.

Figure 4
Median Productivity Ratings for Free Writing (FW) and Structured Writing (SW) Groups Over Six Weeks of instruction

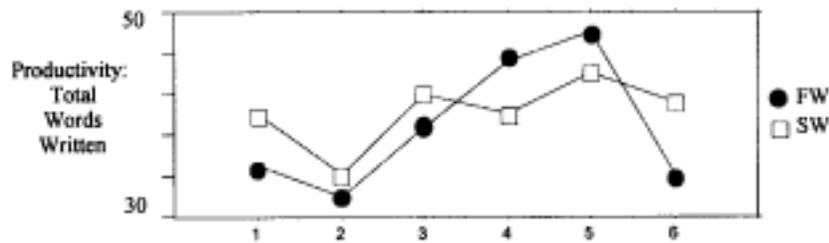


Figure 4 shows considerable variability in writing productivity for both groups. For the FW group, a general increase in words written over weeks one to five dropped greatly in week six. For SW students, a more gradual increase was noted, without the week six drop. The following Table 6 produces the summary statistics for this graph.

Table 6
Trend Tests and Group Differences Between FW and SW Students for Holistic Ratings on a Series of Six Writing Samples

Score	Group	Avg. Gain	Tau	t-test for diffs
Productivity	FW	7.4	.33 (ns)	.89 (ns)
	SW	6.8	.55 (ns)	.58 (ns)

*ns = not significant

Table 6 shows that overall, both groups gained slightly, by about seven additional words written. However, neither group's improvement was significant beyond chance.

Discussion

This study, conducted to judge the relative effectiveness of Process Writing (PW) and Structured Writing (SW) instruction occurred in a six-week summer program for LEP Hispanic students. The study examined four types of writing quality in weekly writing samples from 48 students

(23 FW, 25 SW), randomly assigned to eight classrooms. The non-parametric Tau and 'S' statistics were used to calculate growth over time for each group, and to compare the two groups.

Of the four types of scores derived from student writing samples (Micro-indicators, Analytic ratings, Holistic ratings, Productivity scores), students showed significant growth only in the analytic scores and holistic ratings, and only by SW students. These results were counter-intuitive, as Holistic ratings were considered more sensitive to the FW instruction. Temporarily disregarding statistical significance, the patterns of results from the nine individual scores were mutually confirmatory, supporting SW treatment.

The micro-indicators "percent correctly spelled words" and "percent correct word sequences" showed positive gain for the SW group, and negative gain for the free writing group. Even the largest gain, eleven "percent correct" points, was not statistically significant. At the current rate of improvement, an additional three weeks would be required for statistical significance.

The superior growth of SW students was nearly as pronounced for analytic ratings as for the micro-indicators. For all five Analytic ratings (Topic, Organization, Meaning, Sent., Mechanics), the SW students either surpassed or equaled (for Organization) FW students. Although no between-group differences were significant, SW student gains were statistically significant for all scores but Organization, and FW gains were only significant for Meaning, but no between-group differences were statistically significant.

Results for Holistic ratings similarly were not as predicted. We had hypothesized that benefits of the FW treatment, emphasizing general interpretive feedback through teacher dialogue, would be best measured through a similarly impressionistic (Holistic) evaluation. However, average Holistic gains for SW students were almost four times those for FW students (1.2 versus .3 rating points), although differences did not reach statistical significance. FW students performed relatively better in Productivity (total words rated). Nearly equal, though non-significant gains of about seven words (approximately one word per week) were demonstrated for both groups.

The consistent pattern of superior growth (statistically significant or not) by the SW group causes us to reflect on "common wisdom" versus

the meager research base on LEP student writing. As noted earlier, only two of ten process versus writing studies have involved LEP students (Hillocks, 1986), one favoring free writing, and one showing no differences. The free writing strategy, journal writing, is widely extolled in the literature, but most articles are theory-based treatises, individual case studies, or descriptions of the writing process, with a paucity of student outcome studies. This study is unusual in its adherence to standards of reliable judgment, its emphasis on outcome measures rather than process, and its inclusion of a randomly assigned group of students rather than strategically selected individuals or intact classrooms.

Cautions against assigned writing topics and error correction abound (Graves, 1931; Staton, 1982; Tchudi, 1986; Williams, 1989; Zamel, 1982), warning that students will become bored, intimidated, will suffer undeveloped creativity, will become less fluent (producing less), will suffer "writer's block," and will be unable to apply writing skills to real world situations. We had the opportunity to watch for the occurrence of the first four of these five dangers; they were not observed. But neither did we conduct a dated skill hierarchy "drill-and- practice" approach to structured writing. We implemented not a "straw man," but a reasonable version of structured writing which is widely practiced in general education classes.

Would a longer study have produced different results? By reducing the large number of non-significant findings, certainly "yes"; the power of our tests were restricted by the small number of test probes. However, in those cases where growth trends were negative, or nearly flat (no growth), we would not wish to argue that more time would result in improved results for either group. The question of differential growth patterns between the two treatment groups could reasonably be hypothesized, with more time favoring the free writing group. Staton (1982) found that free writing requires an extended length of time of at least 24 weeks to be most effective. Williams (1989) and Samway (1993) suggest that error correction results in gains for the short run which diminish over time. To answer this question would require a longer study. A second, more tentative answer could be obtained from a curvilinear analysis of the data series, based on nonparametric orthogonal polynomials. We have not yet completed that analysis, partly because of its relatively low power for only six observations.

The findings of this study are not consistent with most current literature on writing instruction with ESL and bilingual populations. However, little of this extant literature is empirical, and fails to offer evidential support for any position. We do not wish to overemphasize or overgeneralize from a single study limited to six weeks of instruction. However, we do wish to strongly urge more such applied outcome studies to support a field governed largely by theoretical and descriptive writings. Such papers, along with advocacy statements, do have an important role in our field, but have somewhat obscured the need for empirical research.

Authors' Note: This study was supported by a Transitional Bilingual grant from the Office of Bilingual Education and Minority Languages Affairs (OBEMLA), USDOE.

References

- Applebee, A. N. (1984). *Contexts for learning to write*. Norwood, NJ: Ablex.
- Barlow, D. H. & Hersen, M. (1984). *Single case experimental designs*. NY: Pergamon Press.
- Darlington, R. B., & Carlson, P. M. (1987). *Behavioral statistics: Logic and methods*. The Free Press, Macmillan.
- DeAvila, E. A., Duncan, S. E., & Navarrette, C. J. (1987). Finding out/descubrimiento. *Teacher's Resource Guide*. Northvale, NJ: Santillana.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219-232.
- Deno, S. L., Marston, D., & Mirkin, P. K. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children*, 48, 368-371.
- Edelsky, C. (1982). *Development of writing in a bilingual program*. Final Report. Volumes 1 and 2. Tempe, Arizona: Arizona State

- Department of Elementary Education (ERIC Document Reproduction Service No. 221057).
- Edelsky, C. (1986). *Writing in a bilingual program: Habia una vez*. Norwood, NJ: Ablex.
- Edelsky, C., & Smith, K. (1989). Is that writing - or are those marks just a figment of your curriculum. In G. Manning & M. Manning (Eds.), *Whole language: Beliefs and practices, K-8* (pp. 183-193). Washington, DC: National Education Association.
- Emig, J. (1971). *The composing process of twelfth graders*. Urbana, IL: The National Council for Teacher Education, (ERIC Document Reproduction No. 058 205).
- Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, 28, 122-128.
- Ferguson, G. A. (1965). *Nonparametric trend analysis: A practical guide for research workers*. Montreal, Quebec: McGill University Press.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. (2Nd ed.). New York: Wiley.
- Florez, V. & Hadaway, N. (1987, January). *Relationship of oral language proficiency and writing behaviors of secondary language learners*. Paper presented at the Southwest Regional Conference of the International Reading Association, Phoenix, AZ.
- Glynn, T., Jerram, H., & Tuck, B. (1986). Writing as an interactive social process. *Behavioral Approaches With Children*, 10, 116-126.
- Graves, D. H. (1981). *A case study observing the development of primary children `5 composing, spelling, and motor behaviors during the writing process*. Final report. NIE Grant No. 678-0174. Durham, NH: University of New Hampshire Press (ERIC Document Reproduction No. 218 653).
- Graves, D. H. (1984). A researcher learns to write. In M. Rose (Ed.), *When a writer can't write* (pp. 1-18). New York: Guilford Press.
- Graves, D. (1987). *Writing: Teachers and children at work*. Portsmouth, NH: Heinemann Publishers.
- Gutstein, S. P. (1983, March). *Using real English: Writing a dialogue journal*. Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages, Toronto, Ontario.

- Hammond, E. R. (1983). *Teaching writing*. New York: McGraw-Hill Book Company.
- Harris, M. (1985). Diagnosing writing process problems: A pedagogical application of speaking-aloud protocol analysis. In M. Rose (Ed.). *When a writer can't write* (pp. 166-181). New York: Guilford Press.
- Hartwell, P. (1984). The writing center and the paradoxes of written-down speech. In G. A. Olson (Ed.), *Writing centers* (pp. 48-61). Urbana, IL: National Council of Teachers of English.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart, & Winston.
- Hendrickson, J. M. (1980). The treatment of error in written work. *Modern Language Journal*, 64, 216-221.
- Hillocks, G. (1986). *Research on written composition*. Urbana, IL: ERIC Clearinghouse on Reading and Communication.
- Huck, S. W., Cormier, W. H., & Bounds, W. G. (1974). *Reading statistics and research*. New York: Harper & Row Publishers.
- Hudelson, S. (1984). Kan yu ret an rayt an ingles: Children become literate in English as a second language. *TESOL Quarterly*, 18, 221-238.
- Hughey, J. B., Wormuth, D. R., Hartfiel, V. F., & Jacobs, H. L. (1983). *Teaching ESL composition: Principles and techniques*. Rowley, Massachusetts: Newbury House.
- Kendall, M. G. (1955). *Rank correlation methods*. London: Charles Griffin & Co., Ltd.
- Krashen, S. D. (1977). The monitor model for adult second language performance. In M. Burt, H. Dulay, & M. Finocchiaro (Eds.). *Viewpoints on English as a second language* (pp. 151 - 156). New York: Regents.
- Kreeft, J. (1983). *Why not really communicate? In dialogue journals?* Washington, DC: Center For Applied Linguistics (ERIC Document Reproduction No. 214 197).
- Mullen, K. A. (1980). Evaluating writing proficiency in ESL. In J. W. Oiler Jr. & K. Perkins (Eds.), *Research in language testing* (pp. 160-170). Rowley, MA: Newbury House.
- Murray, D. M. (1978). Internal revision: A process of discovery. In R. Cooper & L. Odell (Eds.). *Research on composing: Points of*

- departure*. (pp. 85-103). Urbana, IL: National Council for Teacher Education.
- Myers, M. (1983). Approaches to the teaching of composition. In M. Myers & J. Gray (Eds.). *Theory and practice in the teaching of composition: Processing, distancing, and modeling* (pp. 3-43). Urbana, IL: National Council for Teacher Education.
- Omaggio, A. C. (1986). *Teaching language in context*. Boston, MA: Heinle and Heinle.
- Parker, R., Tindal, G., & Hasbrouck, J. (1991). Countable indices of writing quality: Their suitability for screening eligibility decisions. *Exceptionality, 2*, 1 - 17.
- Raimes, A. (1984). Anguish as a second language? Remedies for composition teachers. In S. McKay (Ed.). *Composing in a second language* (pp. 81-96). Rowley, MA: Newbury House.
- Raimes, A. (1986). *Teaching writing: What we know and what we do*. Paper presented at the 20th annual TESOL Convention, Anaheim, CA.
- Samway, K. D. (1993). "This is hard, isn't it?": Children evaluating writing. *TESOL Quarterly, 27*, 233-258.
- Scriven, J., & Glynn, T. (1983). Performance feedback on written tasks for low achieving secondary students. *New Zealand Journal of Educational Studies, 18*, 133-145.
- Siegel, S. (1956). *Nonparametric statistics*. New York: McGraw-Hill.
- Spandel, V., & Stiggins, R. J. (1990). *Creating writers: Linking assessment and writing instruction*. New York: Longman.
- Staton, J. (1982). *Analysis of dialogue journal writing as a communicative event; vol. 1. Final report to the National Institute of Education*. Washington, DC: Center For Applied Linguistics (ERIC Document Reproduction No. 217 327).
- Tchudi, S. N. (1986). *Teaching writing in the content areas: College level*. Washington, DC: National Education Association.
- Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *Journal of Special Education, 23* (2), 169-184.
- Wallace, D. (1987). Dialogue journals: A tool for ESL teaching. In G. Manning & M. Manning (Eds.), *Whole language: Beliefs and practices, K-8* (pp 201-207). Washington, DC: National Education Association (ERIC Document No. 280 316).

- Wienke, J. W. (1981). *Strategies for improving elementary school students' writing skills*. ERIC Document No. ED 209 679.
- Williams, J. D. (1989). *Preparing to teach writing*. Belmont, CA: Wadsworth Publishing Company.
- Yates, J. M. (1983). *Research implications for writing in the content areas*. Washington, DC: National Education Association.
- Young, T. A. (1990). The dialogue journal: Empowering ESL students. *Writing notebook: Creative word processing in the classroom*, 8, 1, 16-17.
- Zamel, V. (1982). Writing: The process of discovering meaning. *TESOL Quarterly*, 16, 195-209.
- Zamel, V. (1987). Recent research on writing pedagogy. *TESOL Quarterly*, 21, 697-715.

Appendix

Analytic Rating Scale

Level	Topic Development
5	Knowledgeable; supported, is well organized; and logical. Clearly a superior paper.
4	Knowledgeable; adequate details and elaboration provided; ideas are relevant to the topic and stated clearly.
3	Adequate knowledge of subject; limited elaboration, limited range of thoughts and considerable lack of details.
2	Most phrases are difficult to understand; often times incoherent.
1	Not enough text to permit evaluation.
Level	Internal Organization
5	Minor problems with the sequencing of thoughts, details, and topic development.
4	Some problem evident with the sequencing of thoughts, details, and topic development.
3	Major problems with the sequencing of thoughts, details, and topic development.
2	Most phrases are difficult to understand; often times incoherent.
1	Not enough text to permit evaluation.

Level	Conveying Meaning
5	Excellent range of vocabulary; conveys meaning accurately; good use of descriptive language.
4	Some use of descriptive language; conveys meaning adequately; words are broad, precise, and literate.
3	Limited and simple word choice; lacks descriptive language usage.
2	Most phrases are difficult to understand; often times incoherent.
1	Not enough text to permit evaluation.
Level	Sentence Construction
5	Creates complex constructions; few errors with parts of speech; little or no evidence of fragmented or run-on sentences; native-like control of grammar.
4	Has minor problems with complex constructions; some errors in agreement, tense number, word/order, function, pronouns, articles, or prepositions; produces some fragments or run-on sentences.
3	Has major problems with complex constructions; produces simple constructions; major problems with fragmented or run-on sentences.
2	Most phrases are difficult to understand; often times incoherent.
1	Not enough text to permit evaluation.
Level	Mechanics
5	Mastery of conventions; very few errors in spelling, capitalization, punctuation, and paragraphing.
4	Some errors in spelling, punctuation, capitalization, and paragraphing.
3	Numerous errors in spelling, punctuation, capitalization, and paragraphing.
2	Most phrases are difficult to understand; often times incoherent.
1	Not enough text to permit evaluation.