Language Learning 50:4, December 2000, pp. 573–616

EFL Business Writing Behaviors in Differing Feedback Environments

Clyde A. Warden
Chaoyang University of Technology

To understand students' reaction to feedback, path analysis was used to generate models of student writing activity. Three groups received differing combinations of feedback and redrafting opportunity over four assignments. A total of 563 first drafts from 141 students were analyzed, resulting in 5,630 data points for modeling. Results suggest that Taiwan business English writing students may implement proofreading and self-directed redrafting behaviors aimed towards improving their writing when objective sentence-level error feedback is supplied with no opportunity for further correction. The addition of redrafting opportunity resulted in generally unproductive writing strategies but also lower objective error rates. Students receiving general teacher-based feedback exhibited redrafting behaviors and a tendency to copy directly from reference material.

What are the steps English as a Foreign Language (EFL) students go through when completing business-oriented English

Clyde A. Warden, Department of Applied Foreign Languages.

The author would like to thank the Overseas Chinese Institute of Technology for hosting this study; Chaoyang University of Technology for support; and the editor, Nick Ellis, and three anonymous reviewers for their insightful comments on earlier drafts of this paper.

Correspondence concerning this article may be sent to Clyde A. Warden, Chaoyang University of Technology, Department of Applied Foreign Languages, 168, Giffeng E. Rd., Wufeng, Taichung County 413, Taiwan, Republic of China. Internet: warden@mail.cyut.edu.tw; World Wide Web, http://www.cyut.edu.tw/~warden/

documents, and can these steps be differentially influenced by the type of feedback provided? Quantitative answers to these questions will help us rationally approach numerous issues in EFL writing instruction. The impact of error correction, for example, can be measured not only in the final product, but also in its influence on students' writing behaviors and strategies. Such issues may be influenced by cultural factors that require a specific examination within the context of the culture in question. For EFL teachers in Asia, the circumstances that surround English instruction often center on instrumental issues, rather than integrative ones, thus raising questions about the appropriateness of wholly adopting imported ESL teaching methods.

Reviewing L2 writing feedback research presents some problems for the EFL context, because most previous L2 studies have been performed on students already located in English-speaking environments, where integration is a clear goal. The question of just which results from L2 research can be applied to the EFL setting is mirrored in the question of what results L2 can draw from L1 research.

The similarity of L1 and L2 writers has been observed by Zamel (1983), who found that skilled L2 writers generally spent more time in editing and revising their work than did unskilled L2 writers and that unskilled L2 writers wrote like unskilled L1 writers. Although similarities have been observed, Raimes (1987) found differences between L1 and L2 writers and suggested that L1 writing instruction methods should not be wholly adopted into the L2 context but adapted to make a better fit. With some caution, we can generally draw on L1 research to help construct a picture of what makes a good writer. Stallard (1974) observed that good writers revised more than poor writers. Sommers (1980) and Faigley and Witte (1981) found that good writers concentrated more on revising content over form. We can see L1 and L2 do have some overlapping findings, although the complexity of L2 should raise some caution when transferring methods. A central question in this transfer is the suitability of responding to students' writing.

Sommers (1982, p. 149) observed teachers' responses to students' writing as "arbitrary and idiosyncratic." This lack of clarity and direction in teacher feedback was also observed by Zamel (1985) in the L2 setting as well as by Brannon and Knoblauch (1982) and Connors and Lunsford (1993). Even if teacher comments are clear, it is not certain that the students, whom the comments are aimed at, can understand such feedback. This is mainly due to the fact that more comments would be given out to weaker students, who are also weak readers (Butler, 1980).

Some researchers in this area have examined the effects of praise and criticism (Cardelle & Corno, 1981) and the emotional content of feedback (Sperling, 1994), whereas others have looked into the goals of comments (Beason, 1993). Truscott's (1996) excellent review of error correction research strongly concludes that correction of errors is counterproductive and cites three studies that found grammar correction to actually be harmful (Kepner, 1991; Semke, 1984; Sheppard, 1992). In the same article, Truscott goes on to observe that people do not like to be corrected and concludes that uncorrected students may go on to be better writers, but this assertion is no more empirically supported than the grammar correction that Truscott is so strongly opposed to. Researchers such as Lalande (1982) and Fathman and Whalley (1990), who have shown benefits from comments on errors, and Liou (1993) and Warden (1995), who found significantly higher error declines for groups receiving specific computer-generated error feedback, remind us that results are mixed. Quantifying the effect of feedback in numerous settings, mixed feedback combinations, and different cultures should be the research goal, including the possibility that error correction is not helpful.

A theory of what error feedback actually does may be a more productive avenue of research than starting out with a direct link between correction and writing quality. Truscott's (1996) review clearly shows the wide range of cultures and situations involved in L2 instruction. Add to this the range of definitions of what good writing is, or even what the goal of a writing class is, and we can begin to understand why empirical results have been mixed (Leki,

1990). It is possible that there exist multiple methods and combinations leading to similar results. An understanding of underlying strategies students are using and exactly how they are affected by different forms of feedback can inform the formation of more useful theory from which teaching methodology can be drawn.

This study constructed detailed models of students' writing steps through the use of computer-based writing assignments in order to examine the impact of differing feedback types not only on the final document, but also on students' writing strategies. Since 1991, I have been developing a software program for Taiwan business writing classes, Quick Business Letters (QBL), which allows detailed tracking of student writing (Chen, 1997; Warden, 1995; Yao & Warden, 1996). The approach includes completion of business writing assignments on floppy disks, outside of class, and submission of files over a network when completed. Editing variables can be measured in addition to numbers and types of errors. I used such a system (Warden, 1995) and found that students receiving computer-generated feedback reduced errors significantly more than students not receiving such objective error feedback but still receiving general comments on overall form and content (a constant variable between the two groups). Reduction of errors was found in subsequent assignments. In view of that finding, the first hypothesis of this study states:

> *H1. Students receiving objective computer-generated error feedback will exhibit writing strategies that prevent errors on subsequent first drafts more than students not receiving objective computer-generated error feedback.*

The QBL system has been successfully integrated into a multi-draft, process-based writing class (Yao and Warden, 1996). Our findings included rapid declines in objective error rates, although the class retained its process-centered emphasis. Thus, the second hypothesis asserts:

> *H2. Students receiving objective computer-generated error feedback and having redraft opportunity will exhibit writing strategies that cause a decline in error rates during subsequent first drafts.*

Chen (1997) used the same software system to measure the amount of text students copied directly from reference material for use in their documents (an approach continued in this experiment). In that study, Chen found that students receiving objective computer-generated error feedback did not exhibit significantly lower objective error rates than students receiving feedback that appeared to be generated by the computer (the correction program actually falsely reported that students' assignments were free of errors). The two groups showed significant differences in editing activity and amount of reference material directly copied. Her conclusion was that students receiving objective error feedback independently undertook editing activities that are similar to those sought in a process-based writing class such as proofreading and redrafting. Although some errors were prevented through these strategies, other errors were introduced. The students not receiving true error feedback reduced errors through more traditional methods (in the Asian context), such as direct copying of reference material into their documents while not employing proofreading or redrafting strategies. Simply examining the final product would not have revealed that although both groups had arrived at the same end, the paths they took were quite different. This develops the third hypothesis:

> *H3. Students receiving objective computer-generated error feedback will exhibit more proofreading and redrafting activity than students not receiving objective computer-generated error feedback.*

Method

Three intact classes of students were selected for this study in order to maximize control of variables while retaining a high number of participants. The class sections were all very similar in nature, in that they were sections of the same academic department, in the same academic year, and completing the same course work.

The study was held at a junior business college in central Taiwan, the Overseas Chinese Institute of Technology. This five-year junior college program includes three years of high school–equivalent work and two years of college-equivalent work. The five-year junior college degree entitles a student, if accepted, to begin university at or above the sophomore year. The majority of students at this school are female, as is the case at other business colleges in Taiwan.

All class sections involved in this study were from the international trade department and in their graduation year (the fifth year of study). Over their five years, international trade students undergo a thorough English program that includes writing, reading, listening, and speaking classes during every semester. English classes center on skill achievement, with little or no exposure to language theory or literature. By graduation, international trade students have had over 42 credits of English classes, amounting to more than 672 in-class hours of English instruction. There is no standardized language exam these students take that can describe their language ability. An informal estimation can be derived from students who have gone on to take the TOEFL exam (5 to 10 percent of students). This department's students often score in the 400 to 520 range (an admittedly wide range, reflecting the fact that non–English majors in Taiwan have large variations in English ability).

Taiwan students remain in the same class section throughout their college careers, meaning that the approximately fifty students who begin school in the same class section will attend most classes together, follow the same schedule, and graduate together. Owing to this structuring, it is difficult to obtain randomly sampled groups for experimentation. All class sections follow the same curriculum and for the most part even have the same teachers throughout their years at college.

Business Writing Class

All three sections were attending English classes. A total of 141 students (see Table 1) participated. Fifth-year required English curriculum included four hours of business English and two hours of English conversation a week. The business English class consisted of two hours a week of international business English writing instruction and two hours a week of practice class, which could include instruction on more general English topics or allow students to work on assignments from the business writing class. Emphasis in the business writing class was on practical skill acquisition that could enable students to produce the numerous written communications often used in Taiwan-based international exporting businesses. Participating sections were taught by a native English speaker and had class instruction in English.

Business Writing Assignments

All three sections were held concurrently during the first (Fall) semester of the 1996–97 school year. Table 2 shows the assignment schedule. All students used the QBL software to complete their assignments. This custom-created program guides students through the creation of Western business letter formats, including headings, openings, and closings (all of which differ radically from standard Chinese business formats). Letter-format accuracy and its correction was consistent across all groups,

Table 1

Students participating in experiment

Class/group	Males	Females	Total
A	4	44	48
B	3	41	44
C	8	41	49

whereas differences in treatment were restricted only to the body of the business letter.

Differences in Treatment Among Groups

Method of feedback and redraft differed, whereas all other variables among the three sections were held constant.

Group A

This group of students had the least complex pattern of homework completion (no redraft), involving the following:

1. Assignment completion and submission over network (including hard copy submission).
2. Computer-generated feedback printed and returned to students.
3. Grades assigned based on number of errors found by the computer, combined with the teacher's score deductions or additions due to formatting, general clarity, and accuracy for the given business letter genre.
4. No redraft was assigned or accepted.

Table 2

Assignment submission dates

Number	Assignment	1st draft due[a]	Redraft due[b]
1	Employment application	October 25	November 11
2	Business trade inquiry	December 8	December 16
3	Response to inquiry for business	December 26	January 1
4	Sales letter	January 10	NA[c]

[a]Groups A, B, and C. [b]Groups B and C only. [c]No final redraft, owing to time constraints.

Group B

The second section was similar to the first, except for the addition of redrafting and delayed grade assignment. Steps for the students included the following:

1. Assignment completion and submission over network (including hard-copy submission).

2. Computer-generated feedback printed and returned to students without the teacher's comments or grades.

3. A redraft completed within a week and submitted (both hard copy and electronic file).

4. Computer-generated feedback printed and returned to students.

5. Grades assigned based on number of errors found by computer combined with the teacher's score deductions or additions due to formatting, general clarity, and accuracy for the given business letter genre (all on redraft only).

Group C

The third section did not receive any computer-generated feedback but instead was given teacher-based non-sentence-level feedback. The process included the following:

1. Assignment completion and submission over network (including hard copy submission).

2. A feedback form attached to the completed first draft. The top half of the feedback form was completed, giving feedback to the student on general areas. No grade was assigned.

3. A redraft completed within a week and submitted (both hard copy and electronic file).

4. The bottom half of the feedback form completed by the teacher and a grade assigned based on overall quality of the paper as well as improvement from the first draft, combined

with the teacher's score deductions or additions due to formatting, general clarity, and accuracy for the given business letter genre.

Feedback

All three groups received some type of feedback after turning in both a printed hard copy of their business letters and an electronic version. The hard copy version of the business letter could contain numerous corrections, from the teacher, written directly on the business letter homework. For groups A and B, computer-generated feedback was supplied on errors within the body of the business letter. This type of feedback was attached behind the original business letter (computer-generated feedback was only supplied in printed form from the teacher). Appendix A shows an example of a student's business letter body and an example of the computer-generated feedback for the student's text. This computer-generated feedback did not contain any other comments from the teacher.

Group C did not receive any form of computer-based feedback but instead received a feedback form. This specific form was adapted from Interactions I (Kirn & Hartmann, 1990), with a few alterations, such as a business letter emphasis, and has been tested and used successfully within a computer-based writing class in previous experiments (Yao & Warden, 1996). The feedback form was split into two identical parts, the top for the first draft and the bottom for the redraft. General topics of content, organization, grammar, and form were covered in the feedback form. For each measure, a box was checked for *Good* or *Needs Work* in the first draft, and *Improved* or *Needs Work* in the redraft (see Appendix B). With this form, students were told that grades were based on improvement and effort between drafts. An area on the right of the form was open for the teacher's comments on what a student should concentrate on during redrafting or for positive feedback on topics that the student did well on. The two-part feedback form was first attached to a student's business letter with the top half

of the feedback form completed. When the redraft was completed, it was attached on top of the first draft and feedback form, stapled together, and all turned in to the teacher again. The hard copy of the business letter submitted to the teacher was returned to the students (for all groups).

Data Types

The first group of variables was chosen to capture students' natural writing behaviors. This required data collection to be noninvasive. The business-letter writing software was on floppy disk, so students were free to complete their assignments wherever and whenever they liked. The measures taken, during the completion of an assignment, were required to be collectible by the computer software with no interference to the writer; objective; easily replicated; and lending themselves to quantitative analysis, yet also reflecting aspects of underlying writing strategies (see Table 3). These variables formed a group labeled *editing* variables and were assumed to give some insight into the amount of editing effort the writer was undertaking; these variables included Editing Time, Keys Added, Keys Deleted, Navigation Keys, QBL Start, Times Printed, and Time to First Print. The QBL software was custom designed to prevent any unusual sources of variation that could confound the intended variables' underlying interpretation. Special design elements included the following:

1. No pasting or cutting features were activated within the interface (this also excluded the ability to paste blocks of text from other text editing programs). The on-disk file was QBL specific and could not be edited by other text editing programs without disturbing the file structure.

2. All variables were reset for the start of each new assignment when the final draft was sent over the network. Networking capability was built into the software.

Table 3

Resulting data types

Variables	Explanation	Interpretation
Total errors	Each student's total errors in a single document: 42 objective error types found by software (see Appendix C)	Increased errors lowers quality of writing product
Book matches	Number of exact phrase and sentence matches with the class textbook examples (allowing for differences in product and company names)	Increased copying from reference material reflects lower amounts of original writing and thus lower levels of editing document
Editing time	Total number of hours spent with the computer program running for a single assignment	Longer time spent with the document reflects more editing activity, as document is examined
Keys added	Total number of keystrokes inputting characters to a single assignment's document (block pasting capability was disabled)	More keystrokes (added or deleted) is a sign of higher levels of editing (with total length of document held constant)
Keys deleted (cut)	Total number of keystrokes cut from a single assignment's document (block deletion capability was disabled)	Higher levels of keys deleted, without matching keys added, shows an emphasis on removing suspected errors in writing
Navigation keys	The number of times any of the up, down, left, right, page up, or page down keys were pressed during a single assignment (mouse navigation capability was disabled)	Movement about the document reflects proofreading of the document
QBL start	Number of times the student's program was started during the completion of a single assignment	Increased opening of the document shows temporal separation of editing efforts; possibly correlated with major revisions of the document
Times printed	Total number of times the document was printed during the completion of a single assignment	This measure is assumed to relate to major revisions to the document, at which time the user prints it out for checking; thus, higher printing rates should relate to higher levels of editing
Time to first print	Total number of hours spent using the student's program from starting a new assignment to the first printout being made	Delaying the first printout shows increased editing of the document, beyond simple input of the required number of words
Words	The total number of words in the body of a student's completed business letter when submitted over network (all assignments required a minimum of 150 words in the business letter's body)	Used as controlling variable for other measurements

3. Navigation was only provided with the up, down, left, and right arrows and the page up and page down keys (see the *navigation keys* variable in Table 3).

4. When a new assignment was begun, the previous assignment was deleted, so students started with an empty file for each first draft. Simultaneously opened files (multiple windows) were not supported.

A second group of three variables was collected after the students' files were submitted over the network. This group included: Total Errors, Book Matches, and Words. Number of words in the body of each business letter was calculated when the files were sent over the local area network; this number was used to control for differences in document length during statistical analysis. Book matches and total errors were calculated through the use of a separate customized software package (QBL TOOLS). A database of sample sentences and phrases from the course textbook was compiled and then used to find the number of exact matches between the textbook and students' documents. Writing errors were found through the same software package that located errors based on the GrammatikV (Reference Software, 1992) error-checking engine, whose database was heavily modified and included over 400 custom errors, decreasing false flagging while raising accuracy in this specific business writing context. This approach was successfully used in Hong Kong by Brock (1990), who modified the Grammatik software package's rule base to find some common errors of Cantonese speakers learning English. Variables were collected in a batch processing approach and all the resulting data saved to a database for statistical testing.

Software Modification

Modification of the Grammatik database began in 1991. Although the software at first appeared inaccurate, especially in its tendency to flag correct structures as errors (false flagging), it immediately appeared useful in the context of English writing

classes in Taiwan. The reasons included large class sizes (normally over forty students, occasionally reaching one hundred), heavy teaching loads (one teacher assigned four or more such classes), and a high rate of repetition of the same types of error among students. Grammatik excelled at handling large batch files with perfect consistency, thus showing the potential to assist in the Taiwan writing classroom setting.

Development goals included improving accuracy of the software for business writing students in Taiwan colleges. Grammatik appears to function in two main stages. Grammatik's Mor-proof parsing engine first parses a sentence into its grammatical parts. This part of the program cannot be modified by the user. Part of this first stage is the determination of fundamental violations of grammar structures, such as missing verbs and incorrect forms. After the first stage of checking, Grammatik compares the contents of a sentence to its database or knowledge-base of rules. These rules can be viewed by the user through the use of Grammatik's Rule Editor program. The rules in this database can be turned on and off through adjustment of Grammatik's options and changed directly through the Rule Editor program.

After extensive testing, it was found that Grammatik's first stage of checking, sentence parsing, could flag false errors. Such flagging of a false error could easily be changed if the rule for the error existed in the database, but rules that were used in the actual parsing were not available for modification. In most cases this problem could be overcome by programming a rule into the second stage of checking that would counter the false flagging of an error in the parsing stage. During the five years of development, new rules were added every semester as common errors were found in students' assignments. These new rules would often result in false errors, because the rule was still under development. For this reason, every error printout was reviewed by a teacher and any false errors crossed out and the error rate manually adjusted before the printout was returned to the student, and the context of the false flagging was noted for ongoing improvement of the system. This served as a final barrier to false errors

and allowed a check on the accuracy of the system. False errors usually ran at 10 percent of total errors. Because these were eliminated by a manual check, the feedback given to students obtained a very low level of false errors. In an effort to minimize false flags, which confused students and reduced overall confidence in the system, the tolerance of the system was raised, resulting in numerous errors not caught by the system. Error rates reported in this study should be interpreted as a subgroup of actual total errors. Internal validity of this study was maintained through the consistent use of the same database across all groups.

Path Analysis

Standardized partial correlation coefficients were generated for all possible interactions of variables in this study. Because ten variables were included, eight variables were held constant and two were tested for correlation, resulting in eighth-order partial correlation coefficients. The level of statistical significance tested for was set at $p < .01$ to assure a robust model (partial correlation coefficients are tested on a t distribution). The partial correlation coefficients generated can be used for connecting together the variables in paths of influence that reveal the relationships among the variables during the completion of the assignments. This technique of path analysis builds a model that can assist in describing relationships (Retherford & Choe, 1993; Sirkin, 1995). By only including in the model the partial correlation coefficients that surpass the minimum significance requirement, a high level of confidence in the model can be obtained, depending on the significance level chosen. It is highly unlikely that all the variables that had influence were measured in this study, and there was no guarantee that the variables measured would connect to form a model at all. If two or more connected variables had no significant correlation with any other variables, the measures would have been orphaned (a clear sign that the research method was missing some important variables, the measurements were not reliable, or

an overall relationship simply did not exist). For this study the resulting models left no orphans; all included variables had lines of influence connecting them to a single model.

During model development, path analysis was chosen over structural equation modeling (SEM), such as LISREL, owing to the initial lack of a model. SEM analysis is confirmatory in nature, whereas this study's model development was exploratory. All four developed path models were later tested for goodness of fit through SEM confirmatory analysis (using SPSS's Amos 3.61 package, 1997). The direction of influence, indicated by arrows, was tested in SEM analysis but should not be interpreted as being causal in nature, because the variables measured were not temporally ordered (they occurred simultaneously rather than in succession). Resulting models included regression coefficients that described the magnitude and direction (negative or positive) of the relationships among the variables and were included in this study to provide an understanding of what differences existed among the three groups.

Results

Analysis centered on the first draft to understand the impact of the differing manipulations on how students approached a new writing assignment and any carryover of behavior from a completed round of draft/feedback/redraft (statistical tests were carried out with SPSS 8.0, 1997). All of the study's variables could be measured across all three groups for the first draft, which was not the case for the redraft, thus allowing for the maximum amount of data. Behaviors during redrafting were examined to check for any deviation from the pattern observed in the first draft. An SEM analysis of three competing models, constructed from the partial correlation coefficients, was undertaken to find which model best represented the data. Resulting models were then used to understand differences among the groups that were related to differences in treatment.

The data did show that the different treatments impacted the groups' writing behaviors differentially in the first draft. A multivariate analysis of variance (MANOVA) test (Wilks' Lambda), including all variables, revealed significant difference due to the assignment ($F_{20, 421} = 3.82, p < .001, \eta^2 = .1$), treatment ($F_{20, 421} = 23.37, p < .001, \eta^2 = .4$), and interaction between the two ($F_{20, 421} = 2.13, p < .001, \eta^2 = .06$). Differences due to assignment were expected, because different business letters present disparate challenges to the students. It is clear that the treatment did play a role, although the specific assignment had an influence.

MANOVA was further employed to reveal specific differences among the groups for each of the variables for the first draft of assignments 2 to 4, followed up by a multiple comparison test with a significance level of $p < .05$. These three assignments occurred after the first feedback was received by students, representing any possible differences among the three groups caused by the treatment. The total errors measure showed a significant difference due to the assignment ($F_{2, 421} = 78.36, p < .001, \eta^2 = .28$) as well as due to the treatment ($F_{2, 421} = 4.99, p < .001, \eta^2 = .02$), while the interaction effect was not significant ($F_{4, 421} = 2.27, ns, \eta^2 = .02$) among the three groups. Descriptive statistics, MANOVA, and multiple-range tests results are summarized in Table 4 and clearly show that the difference in treatment among the three groups impacted the groups differentially.

Groups B and C showed a significant pairwise difference for total errors and keys cut, with Group C having higher means in both measures and Group A falling in the middle. Editing time, navigation, and QBL start all exhibited significant differences between all pairwise comparisons. Group B had the lowest editing time and the least number of times to start the program, while Group C had the highest scores in these measures. Navigation was highest for Group A and lowest for Group B. These results suggest that more editing is undertaken during the first draft when either error feedback is supplied with no redraft opportunity or no error feedback is supplied with redraft opportunity. Groups A and C

Table 4

Means and univariate F *values (sum of squares III)*

	Book matches	Keys added (keystrokes)	Keys cut (keystrokes)	Editing time (hrs)	Navigation (keystrokes)	QBL start	Time to 1st print (hrs)	Times printed	Words in document	Total errors
Pretreatment (all groups)	3.15 (2.18)	1390.28 (584.40)	383.23 (418.12)	1.62 (.99)	1140.94 (1205.68)	5.21 (3.54)	.83 (0.69)	8.57 (8.06)	176.11 (20.49)	9.87 (5.46)
First drafts (2–4)										
Class A (error feedback & no redraft)	4.81_{c} (3.80)	1327.51 (670.02)	300.22 (453.60)	$1.29_{b\ +.71}$ (2.07)	$1143.79_{b\ +.45}$ (1340.43)	$3.36_{b\ +.27}$ (2.69)	.59 (.46)	$4.29_{b\ +.39}$ (4.03)	164.96_{c} (12.88)	7.13 (4.30)
Class B (error feedback & redraft)	5.58 (3.90)	1258.42 (536.00)	204.97_{c} (389.16)	$.83_{ac}$ (.45)	724.44_{ac} (722.95)	2.70_{ac} (2.02)	.44 (.41)	2.83_{a} (3.25)	168.85 (16.26)	6.10_{c} (4.11)
Class C (no error feedback & redraft)	$6.40_{a\ +.37}$ (4.62)	1428.23 (723.18)	$369.20_{b\ +.28}$ (714.22)	$1.70_{b\ +1.12}$ (5.96)	$1005.30_{b\ +.31}$ (1028.93)	$3.54_{b\ +.26}$ (3.91)	.60 (.47)	3.60 (3.84)	$173.37_{a\ +1.85}$ (21.08)	$7.49_{b\ +.30}$ (4.95)
F *values*										
Treatment (2, 421)	9.44**	2.79	3.20*	10.69**	5.85**	2.72	5.33**	4.91**	9.80**	4.99**
Assignment (2, 421)	148.34**	35.39**	4.19*	4.73**	39.17**	3.19*	2.12	1.24	9.28**	78.36**
Treatment × Assignment (4, 421)	3.25*	.62	.09	.45	.30	.43	2.28	1.72	3.63	2.27
Redrafts (2–3)										
Class B (error feedback & redraft)	.13 (.72)	157.37 (207.12)	102.43 (121.10)	.27 (.18)	483.41 (602.32)	2.69 (1.43)	—	2.98 (2.69)	–2.51 (9.66)	–4.88 (4.03)
Class C (no error feedback & redraft)	.08 (1.92)	$386.41_{+.62}$ (449.16)	$244.35_{+.60}$ (292.74)	1.27 (7.63)	$692.75_{+.30}$ (758.67)	2.62 (2.40)	—	3.39 (4.26)	$2.59_{+.39}$ (15.34)	$-.65_{+1.03}$ (2.95)
F *values*										
Treatment (1, 175)	.06	19.02**	16.96**	1.42	4.05*	.06	—	.59	6.91**	80.13**
Assignment (1, 175)	.23	11.87**	1.57	.72	7.53**	.16	—	.84	.39	23.50**
Treatment × Assignment (1, 175)	.16	.53	2.93	.91	5.14*	.19	—	.33	1.16	20.22**

Table 4 (continued)

Means and univariate F *values (sum of squares III)*

	Book matches	Keys added (keystrokes)	Keys cut (keystrokes)	Editing time (hrs)	Navigation (keystrokes)	QBL start	Time to 1st print (hrs)	Times printed	Words in document	Total errors
First Drafts (2–3) + redrafts (2–3)										
Class A (error feedback & no redraft)	3.90_{c} (2.89)	$1476.23_{c-.49}$ (741.96)	$317.74_{c-.54}$ (380.24)	$1.19_{c-.64}$ (.69)	$1181.70_{c-.54}$ (1097.62)	3.65_{bc} (2.61)	—	4.71_{c} (4.09)	164.61_{c} (13.04)	$7.20^{***}{}_{b+1.32,\,c}$ (4.50)
Class B (error feedback & redraft)	4.67 (3.29)	$1509.73_{c-.46}$ (677.62)	$323.91_{c-.49}$ (496.63)	$1.09_{c-.78}$ (.48)	$1292.56_{c-.46}$ (1100.76)	$5.59_{a+.69}$ (2.66)	—	5.57 (3.42)	168.41 (15.23)	$1.17^{***}{}_{c-1.12,\,a}$ (1.59)
Class C (no error feedback & redraft)	$5.04_{a+.32}$ (4.06)	1887.84_{ab} (899.05)	637.26_{ab} (720.59)	1.78_{ab} (1.03)	1900.65_{ab} (1437.44)	$6.10_{a+.68}$ (4.06)	—	$6.96_{a+.49}$ (4.89)	$172.59_{a+.47}$ (19.52)	$6.10^{***}{}_{a-.23,\,b}$ (4.86)
F *values*										
Treatment (2, 270)	6.09**	9.11**	10.24**	21.21**	10.64**	15.46**	—	6.82**	6.02**	90.28**
Assignment (1, 270)	349.19**	49.35**	10.11**	6.84**	48.88**	1.18	—	1.15	18.27**	106.37**
Treatment × assignment (2, 270)	8.38**	.35	.20	1.41	.54	2.42	—	.78	.04	21.58**

* $p < .05$. ** $p < .001$.

Note. Subscripts indicate significant differences, and their effect sizes (Cohen's d), between groups, Tukey-Honestly-Significant-Difference paired comparisons, $p < .05$, where a = Group A, b = Group B, and c = Group C. Standard deviations are in parentheses below means. Degrees of freedom are in parentheses next to F tests.

showed significant pairwise differences for the book matches and words in document measures. Group C displayed higher means in both these measures, with Group B in the center. Number of times printed was significantly different between a high Group A and a low Group B, with Group C in between. In summary, Group B fell at the bottom of six out of eight of the significantly different measures, while Group C had the highest score on six of the measures, and Group A was highest on two measures.

Manipulation Check

While the three groups' treatments differed, namely in redraft opportunity and computer feedback presence, it is possible that students took advantage of a specific treatment regime to maximize grades or minimize workload. Group B may have postponed their effort to the redraft period, because their grades were based on the second draft. Group C may have purposely underperformed on the first draft and placed more effort on the second draft, because their grades were based the amount of improvement between drafts. Four steps were taken to check for such manipulations:

1. Redraft data were collected and compared between Groups B and C.

2. All sentence-level changes made during the redraft stage were quantified and compared between Groups B and C.

3. Sentence-level changes made during the redraft stage were tested for correlation with first-draft error rates.

4. Redraft data (for Groups B and C) were summed with first-draft data and compared among all groups.

Data for this analysis were drawn from the second and third redrafts of Groups B and C and from the first drafts of the same assignments for Group A. These redrafts represent the first redrafts to occur after one completed cycle of treatment and should

reveal any responses to the study's manipulations (the final redraft was not completed in this experiment). Sentence-level changes made during the second draft were quantified through a custom Microsoft Word (1996) macro that compared an individual's first-draft document with the second-draft document and displayed all changes. Any continuous change within a sentence (bounded by periods) was coded as a single change. If one word was cut from a sentence and another word added in its place, the changes were displayed next to each other and coded as one change. If one sentence was removed and another put in its place, the change was also coded as a single change. Within a single sentence, changes separated by unchanged text were coded as distinct changes. While assigning equal value to all changes does not get at the actual effort made, detailed numbers of changes can be seen in the ratio of keys-deleted to keys-added data that were also collected during the redraft stage.

During the first draft Group C showed higher scores on six out of ten measures. This difference is also seen during the redraft stage. Group C, compared with Group B, exhibits significantly higher scores on five measures during the redraft stage. Group C adds an average of 386 keystrokes and deletes 244, whereas Group B adds 157 and deletes 102. Both measures are significantly different at $p < .001$ with $F_{1, 175} = 22.07$ and effect size (Cohen's d) of .62 for keys added, and $F_{1, 175} = 32.71$, effect size of .60 for keys cut. Group C shows higher levels of navigation while redrafting, with navigation-related keystrokes numbering 692.75 compared with Group B's 483.41 ($F_{1, 175} = 4.05$, $p < .05$, and an effect size of .3). The net number of words added to the document during redraft for Group C is positive, mean 2.59, while negative for Group B, mean –2.51, and significantly different at $F_{1, 175} = 6.91$, $p < .001$, and an effect size of .39. The number of mean errors removed during redraft shows that Group B used their feedback to reduce their errors by removing 4.88 errors, whereas Group C removed .65 errors ($F_{1, 175} = 136.05$, $p < .001$, effect size of 1.03). Scores not differing significantly include the editing time, Group B = .27, Group C = 1.27 ($F_{1, 175} = 1.42$, $p = .24$, effect size of .18), the number

of sentence-level changes made to the document, Group B = 5.0, Group C = 5.7 ($F_{1, 175} = 2.02, p = .16$, effect size of .25), number of printouts, Group B = 3.0, Group C = 3.39 ($F_{1, 175} = .59, p = .45$, effect size of .11), and program start, Group B = 2.69, Group C = 2.62 ($F_{1, 175} = .06$, $p = .81$, effect size of .04). While not significantly different at the $p < .05$ level, both editing time and number of sentence-level changes showed some differences between Groups B and C.

While Group B did address their errors, these results showed consistency in behavior across first-draft and redraft stages. Group C continued with their higher levels of keystrokes and time on task. Group B appeared to center on removing the errors pointed out to them in the computer-generated error feedback. This view is further supported by a correlation of .81 ($p < .001$) of errors reported on the feedback and the number of sentence-level changes made during redrafting for Group B, whereas Group C shows a nonsignificant correlation of .17. There was no reason for Group C to address the computer-generated errors specifically, because Group C was never explicitly informed of them. The consistency of behavior between first draft and redraft demonstrates that these two groups did not take advantage of the grading process assigned to them.

Comparisons between Groups A and B showed that in most measures, Group B used the one week of redrafting time to catch up with, but not surpass, Group A. Group B had significantly fewer errors than both Groups A and C ($F_{1, 270} = 90.28$, $p < .001$) and started the program more times than Group A (effect size of .69), but fewer times than Group C (effect size of .68), with a total difference among the groups of $F_{1, 270} = 15.46$, $p < .001$. While postdraft measures such as keystrokes, times printed, and words in document showed the two groups to be equal, we may ask exactly what Group B did during the redrafting stage.

MANOVA comparisons gave a picture of three groups enacting behaviors that resulted in different documents. Group C stood out most clearly, with an emphasis on editing the document in both first-draft and redraft sessions. Even without specific error

feedback, Group C performed somewhat more sentence-level changes than Group B and over twice the keystrokes, with an average of 631 keystrokes to Group B's 260. Objective error rates were higher, and there was an increased tendency to copy material from the textbook when specific error feedback was not present. The differences between Groups A and B, however, are more difficult to tease apart by examining the final results only. Reduction in errors and high correlation of changes with reported errors would suggest that Group B was giving the feedback errors their main attention. This makes perfect sense, as grades were based on the errors remaining in the document, but this does not explain what Group A was doing with their time without redraft opportunity.

Comparisons of means only examine the final result with no understanding of the process. To obtain a better understanding of the underlying strategies being used and how the different variables were influenced by the different treatments, path analysis was employed to build models of the divergent writing strategies used by the three groups. By constructing models, the numerous variables were placed in a context that made underlying relationships within the groups and differences between the groups more apparent.

Prefeedback Model

The prefeedback model was built from all three groups' combined data generated by the first draft of assignment 1. This was before any treatment was implemented, and it supplied a baseline model of what approach students tended to use. Partial correlation coefficients can be seen in Table 5.

The prefeedback model (see Figure 1) showed the general relationships among the variables in this study. Confirmatory SEM analysis obtained a goodness of fit index of .82 ($\chi^2 = 187.16, df = 35$). Of special interest in the prefeedback model was that navigation through the document was related to adding characters, while a

Table 5

Prefeedback partial correlation coefficients (Eighth-order partial correlation, N = *141)*

	Variables	1	2	3	4	5	6	7	8	9	10
1	Book matches	1.00									
2	Keys added	.10	1.00								
3	Keys cut	–.22	.60*	1.00							
4	Editing time (hrs)	.18	.08	.28*	1.00						
5	Navigation	–.07	.42*	.07	.03	1.00					
6	QBL start	.02	.16	–.09	.04	.11	1.00				
7	Time to 1st print (hrs)	–.01	.22*	.00	.74*	.22	.02	1.00			
8	Times printed	–.20	–.07	–.20	.71*	.08	.25*	–.52*	1.00		
9	Words in document	.12	.26*	–.15	.04	–.12	–.05	–.04	–.02	1.00	
10	Total errors	–.31*	.02	–.05	.08	–.13	.08	.02	–.05	.28*	1.00

* $p < .01$.

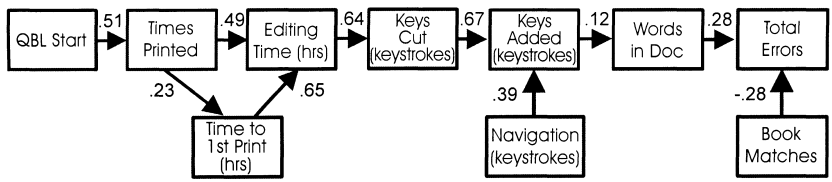

Figure 1. Prefeedback model derived from partial correlation analysis, analyzed with structural equation modeling (results are standardized regression coefficients)

longer time spent working on a document was related to the removal of more characters. This removing of characters was most likely an effort to delete errors found in proofreading. More characters were added with increases in cutting text, which could be the replacement of some of the removed text. There was no link, however, between either of these editing measures and book matches. It seems that the editing was not aimed at, or ineffective at, increasing a document's similarity to textbook examples.

Postfeedback Models

Partial correlation results for each of the three groups (see Tables 6–8) were used to construct corresponding models (see Figure 2). The resulting models were then treated as competing models and analyzed through SEM regression techniques to determine which model best fit the underlying group data. Each model was tested with each group data set, resulting in 9 comparisons (3*3), employing four different measures used for competing models: bootstrapping (Linhart and Zucchini, 1986), the Browne-Cudeck (1989) criterion (BCC), the Akaike (1987) information criterion (AIC), and Bozdogan's (1987) consistent AIC (CAIC). These tests are normally used for comparisons of models, as opposed to GFI, which is useful in comparisons with a baseline model that is assumed to be an accurate representation of the underlying data.

Bootstrapping employs numerous iterations, each of which compares the discrepancy between implied moments from a subsample and the population sample. All the iterations are averaged,

Table 6

Group A postfeedback partial correlation coefficients (Eighth-order partial correlation, N = *144)*

	Variables	1	2	3	4	5	6	7	8	9	10
1	Book matches	1.00									
2	Keys added	–.17	1.00								
3	Keys cut	.24*	.49*	1.00							
4	Editing time (hrs)	–.02	.07	–.04	1.00						
5	Navigation	–.20	.19	.27*	.03	1.00					
6	QBL start	.01	.04	.13	.13	.03	1.00				
7	Time to 1st print (hrs)	.16	.22*	–.03	.04	.10	.05	1.00			
8	Times printed	–.02	–.03	.15	.03	.04	.47*	.22*	1.00		
9	Words in document	.16	.10	–.01	–.03	–.03	.00	–.17	.09	1.00	
10	Total errors	–.44*	.10	.17	.14	–.07	–.15	–.02	–.08	.29*	1.00

* $p < .01$.

Table 7

Group B postfeedback partial correlation coefficients (Eighth-order partial correlation, N = *132)*

	Variables	1	2	3	4	5	6	7	8	9	10
1	Book matches	1.00									
2	Keys added	–.26*	1.00								
3	Keys cut	.25*	.86*	1.00							
4	Editing time (hrs)	.12	–.08	.22	1.00						
5	Navigation	.01	.43*	–.24*	.37*	1.00					
6	QBL start	.04	.24*	–.22	.29*	–.13	1.00				
7	Time to 1st print (hrs)	–.09	–.02	.10	.52*	–.07	0.00	1.00			
8	Times printed	.10	.15	–.28*	.24*	–.12	.17	.18	1.00		
9	Words in document	.27*	.41*	–.37*	0.00	.09	–.16	.02	–.07	1.00	
10	Total errors	–.32*	.06	–.02	0.00	–.06	–.12	–.07	.25*	.31*	1.00

* $p < .01$.

Table 8

Group C postfeedback partial correlation coefficients (Eighth-order partial correlation, N = *146)*

	Variables	1	2	3	4	5	6	7	8	9	10
1	Book matches	1.00									
2	Keys added	–.07	1.00								
3	Keys cut	.07	.62*	1.00							
4	Editing time (hrs)	.08	.67	–.07	1.00						
5	Navigation	.20	.63*	–.08	–.01	1.00					
6	QBL start	.07	.20	.49*	.12	.06	1.00				
7	Time to 1st print (hrs)	.03	.10	–.20	–.05	.14	.04	1.00			
8	Times printed	.09	–.01	.20	.08	.12	.23*	.12	1.00		
9	Words in document	.32*	.26*	–.14	.03	–.01	–.19	–.01	.04	1.00	
10	Total errors	–.35*	–.06	.08	–.001	.08	–.01	.03	–.11	.24*	1.00

* $p < .01$.

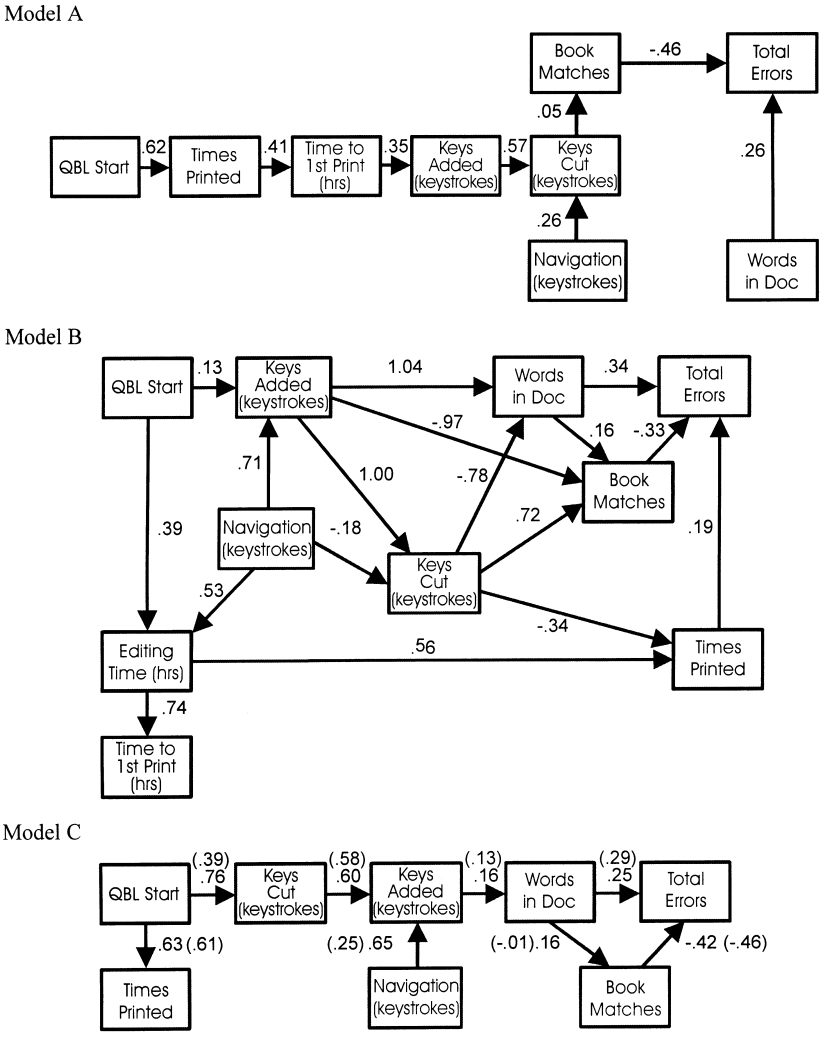

Figure 2. Models derived from partial correlation analysis, analyzed with structural equation modeling (results are standardized regression coefficients with results in parentheses representing Group A's data)

and the model with the lowest discrepancy is indicated as best representing the underlying data. For this study, five hundred iterations were used. The other measures, BCC, AIC, and CAIC, all employ a measure of badness of fit combined with model complexity. Simple models that fit well receive low scores, while complicated, poorly fitting models receive high scores. BCC imposes a greater penalty for model complexity than AIC but less of a penalty than CAIC does.

Results (see Table 9) showed Group A's data were not represented well by the B model and best represented by the C model. Bootstrapping showed Group A's (250.51) model did not represent the data better than the C model (232.14). The remaining three measures all showed model C to be a better representation of the Group A data. This combined with the low chi-squared (84.53) score and the high GFI (.89) pointed to the acceptance of the C model to best represent the Group A data. Group C's data were best represented with the C model, as determined by the lowest bootstrap score (248.16), as well as by obtaining the lowest scores on the remaining three measures.

Group B's data best fit model B and obtained the lowest bootstrap (217.87), AIC (121.43), and BCC (126.38) scores, with the CAIC score (226.27) nearly identical with that of the C model (225.23). The low chi-squared (67.43) and high GFI (.91) also indicated an accurate modeling of Group B's data with the B model. Given these results, it is reasonable both to discuss Groups A and C from the context of the C model, which accurately represented both data sets, and to discuss Group B from the context of the B model.

The C model revealed the underlying strategy employed by Groups A and C. It appeared that these students were placing an emphasis on reviewing and editing their work before submission. If we assumed that a new printout occurred after major revision, or at least every time a student wanted to examine, review, or pass on a letter to classmates for proofreading, then the relationship of starting the program with times printed made sense. Both groups were similar in this relationship. Model C displays a relationship

Table 9

Competing models test results

	Data	Model	Chi-squared	GFI	Bootstrap	AIC	BCC	CAIC
Group A								
(27, 144)	A	A	112.46	.86	250.51 (4.55)	146.46	149.01	213.95
(28, 144)	A	B	137.52	.84	1759.36 (83.01)	191.52	196.02	298.70
(20, 144)	A	C	84.53	.89	232.14 (15.76)	116.53	118.68	180.05
Group B								
(27, 132)	B	A	219.44	.75	343.54 (7.21)	253.44	256.25	319.44
(28, 132)	B	B	67.43	.91	217.87 (7.81)	121.43	126.38	226.27
(20, 132)	B	C	131.11	.84	262.59 (6.40)	163.11	165.47	225.23
Group C								
(27, 146)	C	A	414.32	.69	536.17 (9.37)	448.32	450.83	516.04
(28, 146)	C	B	169.71	.85	8955.50 (538.93)	223.71	228.14	331.27
(20, 146)	C	C	181.51	.80	248.16 (5.81)	213.51	215.62	277.24

Note. Measures other than GFI are for comparison of differing models, where GFI = Goodness of Fit, Bootstrap = mean discrepancy with 500 samples, BCC = Browne-Cudeck criterion, AIC = Akaike information criterion, and CAIC = Bozdogan's consistent AIC. Standard deviations for bootstrap measures appear in parentheses under the bootstrap measures. Degrees of freedom are in parentheses under the group labels.

between starting the program and removing text, which implies that the program was being started with the intention to remove text from a document, which may have already been entered during the first editing session. The adding keys variable corresponded to increases in both keys cut and navigation, showing that increased movement about the document along with increased

removal of text corresponded to increases of material in the document. These relationships appeared to show redrafting taking place. Words in the document increased as material was added to the document; it is important to keep in mind that the material added mediated the keys cut variable that should have decreased document size. On balance, Groups A and C added to the document while moving about the document and removing text. More text in the document gave more opportunity for errors, thus the positive relationship of total errors and words in the document. Errors decreased as matches with the textbook increased, which could have been a strategy for reducing errors.

Group C did appear to use such a strategy, as book matches increased while the words in document also increased. Group A's relationship between words in document and book matches was negative. The editing activities these two groups displayed were not exactly the same. Group C's editing results were more book matches, longer editing time, longer documents, and more errors. Group A's editing tended to be directed at removing possible errors students were aware might exist from receiving previous error feedback. Group A did not navigate about the document so much to add text as to change text, keeping document size down while not entering passages from the textbook or possibly changing passages copied from the book, resulting in fewer book matches.

Although Group B's model appeared complex, there were some important features that stood out. Opening the document was most strongly related to keys added rather than to keys cut. Model B showed that keys were added as navigation increased. Book matches showed quite a complex relationship with four other variables. Increased document size corresponded to more book matches, as was the case with model C; however, keys added showed a strong negative relationship with book matches. It appears that Group B added text that did not match the textbook. Cutting of text showed a negative relationship with words in document and a positive relationship with book matches. This was interpreted as text, once cut, staying out of the document, thereby

raising book matches while decreasing document size. That cutting of text tended to rise as text was added to the document and to fall when navigation increased. Group B did not navigate the document to make corrections through the deletion of text but mostly to add text, which, without further editing, led to increased errors.

Starting the program related directly to an increase in editing time that implied the program was not often started to make quick changes and a printout. As editing time increased, the number of printouts and total errors increased. This supported the assertion that the more work Group B put into the document (timewise), the more errors they had. The link with errors was moderated by keys cut, showing that the removal of text, in an effort to make corrections, can decrease errors. The MANOVA results showed Group B kept errors down not through the use of the removal of material during stages of editing but by minimizing editing time and navigation. Both activities, editing time and navigation, in the B model showed a tendency to increase errors. Rather than concentrating on entering the document for editing out problems, as was the case with Group A, or making changes and adding text that more closely matched the reference material, as is the case with Group C, this group was simply creating the document with no specific overriding emphasis. This makes sense if we assume students were waiting to receive error feedback on which they acted during the redraft stage. The problem with this approach was that the behavior enacted during the first draft stage was certainly not a very productive one.

If supplying specific error feedback and allowing redraft opportunity only led to somewhat equal activity within the document and an emphasis on removing reported errors, we may ask what the advantage was. We expected students would learn how to reduce errors in future writing by finding and correcting previous errors. But this is exactly what we see did not happen with Group B.

Discussion

In this study, the three groups' feedback differed in source—teacher or computer-generated—and redrafting opportunity. All three approaches were applied within a highly similar context. Differences were found in the way the three groups achieved their end results. In the configuration with no redrafting, computer-generated feedback actually encouraged Taiwan EFL students to perform in ways generally thought of as positive, while the slightly different configuration that included redrafting encouraged writing strategies generally seen as negative. Although simple error measurements showed that computer error feedback with redrafting was superior, as seen in Group B's error results, it was the underlying method students used to obtain their error rates that supported this claim.

Group A, with computer feedback and no redraft, directed their efforts at editing the first-draft document. Students in this group took advantage of proofreading and other strategies to catch errors. The assumption that pressing students to avoid errors would lead to shallow editing and an increase in direct copying of reference material was not supported in this study. Group A, being unaware of their specific errors, actually moved about their first drafts making changes and editing the document. In essence, Group A was redrafting on their own before turning in the assignment. The key factor for Group A appeared to be not so much the specific feedback given but the way it was given. Delaying error feedback and not allowing redraft submission caused students to perceive their first drafts not as the first step in the writing process but as the final step in the process. This brought forward redrafting activity that the other two groups delayed until the actual redraft was due.

This activity was enacted by the students on their own. Group A had no redraft opportunity to fall back on if errors were made. The delayed nature of the feedback—error feedback was received along with the grade—meant that it was only an indicator of English writing areas a student was having trouble with. The

feedback could not be used to direct error correction, because only a first draft was accepted.

Opportunity to redraft after receiving very specific error feedback had a negative result on how students approached their first drafts. Students in Group B lacked a direction and may have been waiting until they received the computer-generated feedback to take decisive action. Group B then concentrated on removing the specific errors reported in the feedback. With redrafting time taken up by this activity, and only an average of 16 minutes spent on redrafting, was it possible to undertake any constructive revision of the document? The software was started nearly three times during the redrafting stage, leaving only six minutes of work per session, compared with Group C's average of 29 minutes per session. During revision, Group C exhibited 78% more keystroke activity, compared with Group B (743 mean keystrokes for Group B and 1324 for Group C). All the editing measures for the first draft pointed to Group B performing little revision—editing time, times printed, and QBL start all supported this view. Redrafting effort showed no significant change in behavior. Thus, in view of the writing strategies used, we can conclude that hypothesis *H1* is supported (*students receiving objective computer-generated error feedback will exhibit writing strategies that prevent errors on subsequent first drafts more than students not receiving objective computer-generated error feedback*) whereas hypothesis *H2* is not supported (*students receiving objective computer-generated error feedback and having redraft opportunity will exhibit writing strategies that cause a decline in error rates during subsequent first drafts*). The inclusion of redraft opportunity, while resulting in the lowest error rates of the three groups, led to strategies that simply avoided editing the first draft and emphasized eliminating computer-detected errors during the redrafting stage.

While Group C increased the size of the document by directly copying from the textbook, such copying did not lead to an overall decrease in errors (Group C had the highest error level during the first draft). Indicators of editing activity, such as navigation and times printed, showed that Group C employed a

strategy of making changes to the document during the first draft. Hypothesis *H3* is not supported (*students receiving objective computer-generated error feedback will exhibit more proofreading and redrafting activity than students not receiving objective computer-generated error feedback*). Although the inclusion of objective error feedback did activate editing-like activity in Group A students, it failed to do so in Group B, apparently owing to the inclusion of redrafting. Previous studies using QBL have compared computer-generated error feedback to no feedback or modified forms of feedback. A more realistic comparison shows that teacher-based non-sentence-level feedback can also have positive results. Although the increased dependence on copying reference material is not desirable, students do undertake editing activities.

Group C's teacher-based approach had the least impact on the prefeedback model and increased the importance of copying from reference material. Years of education emphasizing an objective approach may have left these students uncertain how to act in a self-directed, process-oriented activity. Taiwan EFL students may not have a wide-ranging language background and exposure from which to jump into action confidently. In view of Group A's results, these EFL students did have the ability to activate editing behaviors that not only improved the document through revision, but also avoided the traditional strategy of dependence on copying reference material. It may simply be that the motivation these students best understood was related to the traditional objective error-discovery methods so common in Asian classrooms across all fields of study.

With these results, we may begin to understand why previous research results on error correction have been mixed. If we had only examined error rates, the conclusion would be that Group B exhibited superior performance. Including the examination of redrafting would reinforce such a conclusion because Group B effectively zeroed in on their errors. If we had concentrated on editing activity, the conclusion would be that Group C was superior. A more inclusive examination has revealed that specific error feedback combined with redraft opportunity may result in the

worst of both worlds. In Group B's results, may we be witnessing some of the negative impact of error correction that Truscott (1996) observed from previous studies? Here students are told what they did wrong, why it is wrong, and what to do about it. They are then given a chance to act on that feedback in a redrafting stage. The result for Group B was lower editing activity. Especially worrying were the relatively low levels of editing time, navigation, times printed, and the apparent emphasis on homing in on and removing reported errors. It would be difficult to argue that less time spent on the document, with less movement and fewer draft printouts, can improve the document. The inclusion of error feedback without redrafting (Group A) may trigger a critical view of error production and activate previously learned strategies along with a self-awareness that all combine in more productive proofreading and editing activities.

Implications

EFL instructors presently have little choice but to direct their teaching efforts to match imported methodologies, even though little is understood about the actual writing behavior of their own students. This approach may not align with the reality created by social, cultural, and historical trends in which EFL students have been, and continue to be, immersed. Dispositions of specific national educational traditions in Asia may not easily produce students open to an approach that makes assumptions about motivations, since the foundations of such motivations may differ. In Taiwan, for example, entrance exams begin for the high-school level and continue through graduate school with many nationally unified (standardized) exams. These exams always include a test of English with an emphasis on vocabulary and grammar accuracy.

This does not dismiss the importance of implementing a multiple-stage process when writing. The question is how students can learn that writing is a process and then actually incorporate that understanding into their writing behavior. Simply promising that the end result will be improved writing quality may work for

language majors at the university level but is difficult to implement in a normal Taiwan EFL classroom. We have found direct application of multiple drafts and non-sentence-level feedback may obtain a lower level of self-directed revision and increased dependence on reference material when compared with objective feedback with no required redrafting. In other words, Taiwan EFL students may respond best to what they understand most. Once shown that specific errors exist in their documents, these students will expend energy to create better documents. Such a conclusion, by its nature, is not universal but certainly may be important in similar contexts where English writing is taught to students from educational backgrounds that emphasize memorization, repetition, and objective standards.

Caution is called for in making claims on the external validity of these findings for a number of reasons. First, the cultural context of this study, Chinese in the Republic of China, may play a central role in the results, because the education system places a heavy emphasis on product quality. This may explain why students respond to non-sentence-level feedback by copying directly from the textbook. Such an emphasis is common throughout Asia, but this may differ radically in other regions. Second, because the students in this study were not English majors, their English skill varied widely. It is common in Asia to find English classes with advanced and basic students grouped together. More or less variation in skill levels may impact results. Third, the software employed was purposely disabled so that confounding variables could be controlled, that is, no cutting and pasting features. This restriction was a compromise that must be taken into account when attempting to apply the results to a wider context. Within the framework of this study, this unusual feature did not raise any questions on the part of students. The software was presented as specifically designed for the creation of business letters and as such actually assisted students in creating the headings, openings, and closing of their business letters. Software use was not restricted to specific locations or times, thus increasing the opportunity to capture students' natural writing behav-

iors. These factors, while conducive to a research project, may not accurately reflect all variables in other classroom settings.

Finally, we must take care in making any assumptions about which group actually had preferable behavior. What makes these findings interesting is that students developed behaviors that have not been traditionally associated with the specific type of feedback they received. Such quantification can help us make more informed decisions in the classroom. For example, the present study raises the question of immediate error feedback: If students are given immediate feedback, rather than delayed, what is the result? From the current study it would appear that specific error feedback, with opportunity to use it during correction, leads to less emphasis on proofreading-like activities, although objective errors are reduced. Thus the benefit of such an approach can only be judged within the context of what the goal is—objective error reduction (product), independent editing strategies (process), or some combination of both. It is only through such understanding of underlying behaviors and complex relationships, as well as instructional goals, that EFL teachers can create a more informed pedagogy.

Revised version accepted 07 January 2000

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Beason, L. (1993). Feedback and revision in writing across the curriculum classes. *Research in the Teaching of English, 27,* 395–421.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52,* 345–370.

Brannon, L., & Knoblauch, C. H. (1982). On students' rights to their own texts: A model of teacher response. *College Composition and Communication, 33,* 157–166.

Brock, M. N. (1990). Customizing a computerized text analyzer for ESL writers: Cost versus gain. *CALICO Journal, 8,* 51–60.

Browne, M. W., & Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research, 24,* 445–455.

Butler, J. (1980). Remedial writers: The teacher's job as corrector of papers. *College Composition and Communication, 31,* 270–277.

Cardelle, M., & Corno, L. (1981). Effects on second language learning of variations in written feedback on homework assignments. *TESOL Quarterly, 15,* 251–261.

Chen, J. (1997). Computer generated error feedback and writing process: A link. *TESL-EJ* [On-line], *2* (3). Available: http://www-writing.berkeley.edu/TESL-EJ/ej07/a1.html

Connors, R., & Lunsford, A. (1993). Teachers' rhetorical comments on student papers. *College Composition and Communication, 44,* 200–223.

Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication, 32,* 400–414.

Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178–190). New York: Cambridge University Press.

Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *Modern Language Journal, 75,* 305–313.

Kirn, E., & Hartmann, P. (1990). *Interactions I: A writing skills book.* New York: McGraw-Hill Publishing Company.

Lalande, J., II. (1982). Reducing composition errors: An experiment. *Modern Language Journal, 66,* 140–149.

Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 57–68). Cambridge: Cambridge University Press.

Linhart, H., & Zucchini, W. (1986). *Model selection.* New York: Wiley.

Liou, H. (1993). Investigation of using text-critiquing programs in a process-oriented writing class. *CALICO Journal, 10* (4), 17–38.

Microsoft Word 97 [Computer software]. (1996). Redmond, WA: Microsoft Corporation.

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL college student writers. *Language Learning, 37,* 439–468.

Reference Software International [Computer software manual]. (1992). *Grammatik 5, user's guide.* San Francisco: Reference Software International.

Retherford, R. D., & Choe, M. K. (1993). *Statistical models for causal analysis* (pp. 93–118). New York: John Wiley & Sons, Inc.

Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals, 17,* 195–202.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal, 23,* 103–110.

Sirkin, M. R. (1995). *Statistics for the social sciences* (pp. 436–442). London: Sage Publications.

Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication, 31,* 378–388.

Sommers, N. (1982). Responding to student writing. *College Composition and Communication, 33,* 148–156.

Sperling, M. (1994). Constructing the perspective of teacher-as-reader: A framework for studying response to student writing. *Research in the Teaching of English, 28,* 175–207.

SPSS Amos 3.61 [Computer software]. (1997). Chicago: Small Waters Corporation.

SPSS 8.0 [Computer software]. (1997). Chicago: SPSS, Inc.

Stallard, C. K. (1974). An analysis of the writing behavior of good student writers. *Research in the Teaching of English, 8,* 206–218.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46,* 327–369.

Warden, C. (1995). Expert system impact on writing errors in Taiwanese business English classes. *CAELL Journal, 6* (2), 22–29.

Yao, Y., & Warden, C. (1996). Process writing and computer correction: Happy wedding or shotgun marriage? *CALL Electronic Journal* [On-line], *1* (1). Available at http://www.lerc.ritsumei.ac.jp/callej/1-1/Warden1.html

Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly, 17,* 165–187.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19,* 79–102.

Appendix A: Feedback Example and Note on Accuracy

I resently is graduated from Feng-Chia college, and have ever listened to your speech about finding a applicable job for a graduate on Ouctober 23 in department store. I hope to talk to you about a applicable sales position.

I am presently looking for a position after graduation from international trade department.I took part-time work in bookstory department of Chung Yo Department store. I already work in a sale department for three years. So the work I now does includes a assistant saler. My characteristics is the ability to persevere, and play computer.

I think my experence over the past three years,and studying in college has made me have the ability to become a responsible saler.The reason I took part-time work in department is that I hope to got a position in a overseas company to realize what I learned in school.

For more information,please contact me at my home on Sundays or every evening after 8:00 pm., at (04)2555376. I look forward to hearing from you.

======

Check: resently
Rule Class: Spelling
Advice: Spelling error.

======

Check: is graduated
Rule Class: Custom Rule Class 1
Advice: 'Graduate' is a verb and does not require a verb before it.

======

Check: have ever
Rule Class: Custom Rule Class 1
Advice: awkward construction; cut 'ever'

======

Check: Ouctober
Rule Class: Spelling
Advice: Spelling error.

======

Check: a applicable
Rule Class: Article
Advice: You may want to use 'an' before 'applicable'.

======

Check: department
Rule Class: Noun Phrase
Advice: An _article_ or other _modifier_ usually precedes the word 'department'.

======

Check: .
Rule Class: Punctuation
Advice: Usually this punctuation mark ('.') needs a space after it.

======

Check: I
Rule Class: Sentence Variety
Advice: You have used 'I' to begin the last 3 sentences. Consider more variety with your sentence openers.

======

Check: bookstory
Rule Class: Spelling
Advice: Spelling error.

======

Check: bookstory department
Rule Class: Noun Phrase
Advice: An _article_ or other _modifier_ usually precedes the word 'department'. Check also for missing hyphenation.

======

Check: I has
Rule Class: Subject-Verb Agreement
Advice: The _subject_ 'I' requires a plural verb, not the singular verb 'has'.

======

Check: So
Rule Class: Custom Rule Class 2
Advice: Do not begin a sentence with the word 'So,' which is usually used to introduce a clause or phrase.

======

Check: a assistant
Rule Class: Article
Advice: You may want to use 'an' before 'assistant'.

======

Check: work I now does
Rule Class: Subject-Verb Agreement
Advice: The _subject_ 'I' requires a plural verb, not the singular verb 'does'.

======

Check: saler
Rule Class: Custom Rule Class 2
Advice: Do you mean 'seller' or 'salesperson'? A person who sells or markets products or services.

======

Check: characteristics is
Rule Class: Subject-Verb Agreement
Advice: The plural _subject_ 'characteristics' requires a plural _verb_, not the singular verb 'is'.

======

Check: play computer
Rule Class: Custom Rule Class 1
Advice: you do not 'play' with a computer; rewrite to include: use, operate, type on, program, study, etc.

======

Check: think
Rule Class: Custom Rule Class 2
Advice: business writing should not include what you 'think' and you should not tell the reader what to think; replace with: am confident, am sure, am convinced, etc.

======

Check: experence
Rule Class: Spelling
Advice: Spelling error

======

Check: ,
Rule Class: Punctuation
Advice: Usually this punctuation mark (',') needs a space after it.

======

Check: saler
Rule Class: Custom Rule Class 2
Advice: Do you mean 'seller' or 'salesperson'? A person who sells or markets products or services.

======

Check: .
Rule Class: Punctuation
Advice: Usually this punctuation mark ('.') needs a space after it.

======

Check: to got
Rule Class: Infinitive
Advice: The context suggests that you have either an _infinitive_ error or a _homonym_ error.

======

Check: a overseas
Rule Class: Article
Advice: You may want to use 'an' before 'overseas'.

======

Check: ,
Rule Class: Punctuation
Advice: Usually this punctuation mark (',') needs a space after it.

======

Check: pm
Rule Class: Spelling
Advice: Spelling capitalization error.

Note. All feedback pages were screened by the teacher for false errors before being returned to students. False errors were manually crossed out in screening and the total number of errors adjusted.

Appendix B: Process Feedback Example

First Draft	Good	Needs work	Comments
Content			
1. Information is clear and to the point.			Try to center on the product's most useful features.
2. Letter has one central idea (focus)			Remember that this letter should sell the product, not just answer questions. Stay focused on making the sale.
Organization			
1. Unity (each paragraph has one basic purpose)			Each paragraph has a clear topic.
2. Coherence (sentences are connected building up the main idea)			Some sentences are repeating the same product feature. Each sentence should contribute to the overall advantage of the product; don't repeat.
3. Continuity (the connection of the sentences is smooth)			Avoid using conjunctions to start each sentence. Try to have each sentence build on and extend the previous sentence's idea.
Grammar			
1. Choice of words			Business writing often requires a bit more formal vocabulary. Avoid making the letter like a conversation.
2. Use of business vocabulary			Good use of vocabulary for the product's feature set.
Form			
1. Paragraph form			Good paragraph structure. Try not to repeat ideas within a single paragraph.
2. Spelling			Be careful with typing errors; maybe get a classmate to help you proofread.
3. Use of punctuation			Look at the use of spaces with English punctuation.
4. Business letter format			Choose one style and stick with it, like the block style. Do not mix styles.

Note. The form was repeated for the second draft feedback.

Appendix C: Error Types Checked by Software

Error type	Percentage of occurrence in experiment pre-test & assignments 2–4, first drafts				Error example
	Pre	A 2–4	B 2–4	C 2–4	
Abbreviation	.44	.00	.00	.10	Mr. Smith Ph.d. from Zion CO. will arrive next week.
Adjective	.80	.79	1.03	.68	This is an interested story.
Adverb	.66	1.19	.13	.77	She certain is smart, but she is also stubborn.
Article	1.61	.59	1.16	.68	We sell our products in North American Market.
Capitalization	2.12	14.55	11.90	7.05	Tomorrow, i want to visit Bill.
Clause	.00	.00	.00	.00	James went to the tennis match. Even though it was raining.
Colloquial (not appropriate in business writing)	.07	.00	.00	.10	The director will make a decision when he is good and ready.
Comma splice	.44	1.09	1.03	1.16	He smokes when he is working overtime, it keeps him awake.
Conjunction	.07	.00	.00	.00	We had to choose between English or French.
Custom*	20.70	20.69	22.64	20.39	Go in and open the light. We have ever bought your products before.
Incomplete sentence	2.49	.89	1.81	1.84	Our wonderful president who devoted many years of service.
Jargon	.29	.00	.26	.00	Let us interface next week over lunch.
Long sentence	1.32	.69	.52	1.06	(Set at 30 words)
Noun phrase	6.51	10.10	10.61	10.05	He drove motorcycle
Possessive form	1.17	3.66	3.62	7.15	The secretarys desk was covered with work.
Preposition	.00	.00	.26	.10	Everyone in our office must comply to the new regulations.
Pronoun case	.29	.30	.39	1.16	Everyone has their own goal.
Pronoun number	.37	1.49	1.55	.77	They was going to the fair.

Appendix C (continued)

Error type	Percentage of occurrence in experiment pre-test & assignments 2–4, first drafts				Error example
	Pre	A 2–4	B 2–4	C 2–4	
Punctuation	9.58	1.58	1.81	4.93	Tomorrow after lunch we can study,
Questionable usage	.07	.10	.00	.00	The vitamins were a preventative measure.
Redundant	.07	.00	.00	.00	Once you use a computer, you will never revert back to using a typewriter.
Relative pronoun	.00	.10	.00	.19	Her green coat, that she bought in February, has a tear.
Repeated words or punctuation	.22	.10	.65	.10	We all like to travel to to Canada.
Sentence variety	12.73	2.87	2.46	4.35	I would like you to ship before June 20. I could open a letter of credit in your favor within the week. I will wait for your decision.
Similar words	.07	.30	.26	.68	The sentences sited are from Shakespeare.
Spelling	28.02	22.38	21.47	20.48	The acter, who is a techer, had the leading part.
Subject/verb agreement	3.88	9.70	9.44	10.05	My mother always encourage me.
Verb form	1.32	1.58	1.42	2.61	She catching the bus.
Verb object	2.34	1.68	2.07	1.55	She fixed up.
Wordy	1.68	2.18	2.20	1.45	In any case, we can give you a low price.
Total	≈100	≈100	≈100	≈100	

Note. Errors checked by the software but not occurring in this study are not included in this table.
* These errors are the expanded database, including common errors of Chinese students (containing over 400 rules).