Effects of the Red Pen

Harriet D. Semke
Westmar College

ABSTRACT A research study, conducted for ten weeks with 141 first-year, third-quarter German students at the University of Minnesota compared the effects of four methods of teacher treatment of free-writing assignments: 1) writing comments and questions rather than corrections; 2) marking all errors and supplying the correct forms; 3) combining positive comments and corrections; and 4) indicating errors by means of a code and requiring students to find corrections and then rewrite the assignment.

Results of this study indicate that student progress is enhanced by writing practice alone. Corrections do not increase writing accuracy, writing fluency, or general language proficiency, and they may have a negative effect on student attitudes, especially when students must make corrections by themselves.

A traditional assumption has been that written work in foreign language classes must be corrected meticulously. If it were not corrected, errors would be deeply ingrained, and it would be doubly difficult for learners to use the language correctly. However, in spite of corrections, teachers have found that students continue to repeat the same mistakes.

Correcting students' free-writing is a tedious task for the teacher. In fact, the amount of free-writing assigned often may be determined more by the amount of time a teacher has to correct it than by the amount believed to be most beneficial to a student's learning.

Harriet D. Semeke (Ph.D., University of Minnesota) is Associate Professor of German at Westmar College; Le Mars, IA.

Not only is the task of carefully correcting written work time-consuming, but the immediate feedback a teacher receives is frequently negative. The return of papers covered with the inevitable red marks results in looks of disappointment and discouragement on students' faces. The teacher wonders if the students will even bother to read the corrections, to say nothing of learning from them. The likelihood that papers will end up in the waste basket after having received only a cursory glance is verified by an experiment conducted by Marzano and Arthur, who found that the hours teachers spend correcting students' free-writing is an exercise in futility.[1]

The teacher may also be frustrated because of the realization that the number of red marks on papers do not adequately reflect the quality of the students' work. Many of the mistakes may not actually interfere with communication, but the red ink causes them to loom so large that the learner is unaware of all the good, or at least comprehensible, language which has been produced. As a result, students do not feel the affirmation and positive reinforcement that their efforts have merited.

Current educational literature points up the relationship between affective and cognitive factors in the learning process. As just one example, research conducted by Talmage and Eash indicates that student achievement is closely related to student attitude.[2] Anything which has a negative effect on attitude tends to retard learning.

Since free-writing assignments often result only in teacher stress and student disappointment, both of which may have negative effects on progress in language acquisition, the temptation may be to

forget free-writing assignments completely. Yet many foreign language methodologists affirm that language skills are interrelated and that writing reinforces the other basic skills of listening, speaking, and reading. Writing, as a productive skill, is most closely related to speaking, and, in contrast to speaking, is something which students can do alone, as homework. Free-writing is therefore an important means of maximizing the amount of practice students have in producing language. The problem lies not in the writing activity itself, but in the teacher response to the student effort.

This leads to the following question: what are the alternatives to the typical method of correcting each error? Some methodologists advocate a system of simply marking places where errors occur, using a code to indicate the type of error, and requiring students to correct their own work.

Another method of treatment which has been suggested, particularly by teachers of English like Kelly[3] and Wagner,[4] and by teachers of English as a second language like Burt,[5] is to make no corrections, but to respond to the content with written comments and questions, indicating that the message has been understood. Kelly states that the teacher's first responsibility is to listen to what students are saying when they speak. Instead of making red marks, she finds it is more helpful to respond and question, making inquiries which will keep students "talking on paper." Her position is that later, after students realize that they are communicating and someone is understanding the message, they will be receptive to suggestions and even seek advice on how to improve the mechanics of their writing. Research by Rinderer[6] supports the theory that, at least in the teaching of English, a teacher's written, supportive comments have a positive effect on students' motivation toward writing improvement, while corrections tend to stifle motivation.

Problem

In order to obtain some concrete evidence regarding the effects of various methods of treating students' free-writing assignments in foreign language courses, a research project was carried out during the spring quarter, 1980, in the German Department at the University of Minnesota. The basic questions investigated were:

1. Does correcting errors on students' free-writing assignments facilitate students' ability to write and increase students' total language competence, or is correction detrimental to students' progress in language acquisition?
2. Does giving supportive comments have a positive effect on student attitudes?
3. If corrections are necessary for efficient learning, does the addition of reinforcing comments counteract any negative effects on students' attitudes which the corrections are apt to produce?
4. Do students achieve more when they are forced to correct their own mistakes?

Methods and Procedures

Subjects were 141 students[7] enrolled in the third quarter of first-year German. These students were divided among eight sections, each of which was taught by a different instructor. The course met 5 days per week for 10 weeks.

For the purpose of the experiment, a weekly free-writing assignment in the form of a *Tagebuch,* or journal, was incorporated into the course. This assignment permitted students freedom to write about anything they wished—they could recount their activities for the week, or they could discuss a topic of their own choosing.

Of the eight sections of the course, two sections were assigned to each of four experimental groups. (See Table 1.) Since course content (as well as tests) was identical in all sections, differentiation among the experimental groups was only on the basis of method of treating the students' free-writing assignments. This was as follows:

Table 1

Experimental Groups	N
Group 1—comments only	46
Group 2—corrections only	27
Group 3—corrections with comments	38
Group 4—student correction	30
Total N	141

On the *Tagebuch* assignments of group 1 no errors were marked. Instead, students received comments and questions in response to the content of the writing. Grades were based solely on the *amount* of understandable German produced. Since the students had had essentially no free-writing experience, during the first two weeks 100 words were required for an A. For the next two weeks 125 words were required for an A, and thereafter it took 200 words to merit an A; a B required 150 words, a C 100 words, and anything less than 100 words was unacceptable. Any incomprehensible language was bracketed and did not count as part of the total number of words.

The *Tagebuch* assignments of group 2 were corrected with correct forms written in. Group 3 was treated with a combination of the above methods; that is, errors were corrected and comments in response to the content were added. On the writing of group 4, errors were marked by means of a symbolic code. These students then attempted to cor-

rect their own errors and rewrote the assignment the following week. On the second "corrected" versions, group 4's errors were marked with correct forms supplied, in the same manner as for groups 2 and 3. Groups 2, 3, and 4 began with a 50-word minimum and worked up to a 100-word minimum after the fourth week. The grades for these groups were based on a ratio of mistakes to the number of words written. An A required at least 90% accuracy, a B at least 80%, a C 70%, D 60%, and anything below that was an F. During the 10-week quarter, a total of nine compositions were written. Since group 4 did the rewrite every second week, this group generated only slightly more than one-half as much new material as the other groups.

All groups were given pretests and post-tests. These consisted of a timed, free-writing sample and a timed, multiple-choice cloze test which yielded the following three measures of competence in German:

Free-writing	—accuracy
Free-writing	—fluency
General language	—proficiency

The directions for the free-writing test read as follows:

> You will have ten minutes to write as much as you can in German. Just write whatever comes to your mind. Use complete sentences and write in paragraphs, but the paragraphs do not have to relate to each other. You may write on as many different topics as you wish. Write as accurately as you can, but the primary emphasis is on the *amount* of intelligible communication in German that you can produce in this limited time.
>
> Do not begin until the teacher tells you. When the time is up, stop immediately, even if you are in the middle of a sentence.

This test was chosen in keeping with Brütsch,[8] whose research indicates that an actual writing sample is the best measure of writing ability. Scoring was patterned on the procedure used by Brière: a) the ratio of errors to words indicated the level of *accuracy,* and b) the number of words produced in the 10 minutes yielded the score for fluency.[9]

The cloze test consisted of 57 multiple-choice items constructed according to procedures used by Carstens.[10] The directions and a sample follow:

> In the attached reading selections, a number of words have been omitted. Select from the words in parentheses the one which *best* fits both the grammatical structure and the context of the sentence and paragraph. *Underline* this word. *Do not write in* the words.
>
> You will have exactly 10 minutes to do the test. While you are not expected to be able to complete the entire test in that time, work as rapidly as possible and do as much as you can. There is no penalty for guessing.
>
> Sample item from the cloze test:
>
> Der Schuhverkäufer wurde ___(1)___ ärgerlich.
> (neue, langsam, glaube, und)
> Er hatte seiner ___(2)___ schon mindestens zwei
> (nicht, sie, einen, Kundin)
> Dutzend ___(3)___ gezeigt und sie konnte sich
> (sagte, war, Schuhe, glaube)
> ___(4)___ entscheiden.
> (der, nicht, zwei, wurde)
>
> Cloze test translation:
>
> The shoe salesman became ___(1)___ angry. He
> (new, slowly, believe, and)
> had shown his ___(2)___ already at least
> (not, she, a, customer)
> two dozen ___(3)___ and she could
> (said, was, shoes, believe)
> ___(4)___ decide.
> (the, not, two, became)

Oller[11] and Shohamy[12] cite much data supporting cloze tests as efficient, valid, and reliable measures of global language proficiency. They conclude that results indicate competence rather than mere performance. Carstens found that the multiple-choice cloze test is not only less time-consuming and less frustrating to the student, but also makes a greater distinction among levels of language learning than does the normal cloze test. For purposes of validation this test was administered to three native speakers of German, all of whom completed it in less than five minutes with perfect scores.

The same tests, with the same testing procedures being used, were administered both as pretests and post-tests. In addition, a background questionnaire was completed by each student at the time of pretesting, and an attitude questionnaire was used at the time of post-testing.

The background questionnaire showed that the treatment groups were homogeneous in terms of background in German. An interesting difference, however, was that a larger percentage of group 2 (corrections only) was planning to continue the study of German beyond the course 1-103. (See Table 2.)

Table 2

Percentage Planning to Continue German Beyond 1-103	
Group 1 (comments only)	50%
Group 2 (corrections only)	63%
Group 3 (corrections with comments)	42%
Group 4 (student corrections)	50%

A native speaker of German, who also had had experience in teaching German 1-103 at the university, but who was not currently an instructor, corrected and scored the pretest and post-test free-writing samples, as well as all of the weekly free-writing assignments. This procedure was spot-checked by the researcher to be sure that error marking was uniform. All comments, both for group 1 (comments only) and group 3 (corrections with comments) were written by the researcher. This assured consistency of treatment among the course sections. A minimum standard of giving at least two full-sentence responses, one of which was usually a question, was observed.

Besides conveying to the students the message that their German was understood and encouraging them to write more, an additional purpose of the comments was to model correct usage. Whenever possible, without being unnatural, items or forms which the student had used incorrectly were included in the responses.

Students counted their own words on the weekly *Tagebuch* assignments. Spot-checks indicated that these were nearly always accurate.

Data Analysis

The test results were analyzed using an analysis of covariance, a statistical technique which adjusts post-test scores for initial differences, as demonstrated by pretest scores, thus permitting an examination of the mean achievement of the experimental groups.

The analysis of covariance was followed by a set of completely independent contrasts, which made the following three comparisons:

1. Group 1 with groups 2, 3, 4 (no correction vs. correction)
2. Group 4 with groups 2 and 3 (student corrections vs. teacher corrections)
3. Group 3 with group 2 (correction with comments vs. correction only)

These comparisons were tested for significant differences between adjusted post-test means.

Results

The results are presented in Tables 3, 4, and 5, which show the adjusted post-test means of each group. The statistical analysis showed that there was not significant difference among treatment groups in terms of writing accuracy. (See Table 3.)

Table 3
Results of Writing Accuracy Test

Treatment Group	Adjusted Mean
1 (comments only)	74.88
2 (corrections only)	78.44
3 (corrections with comments)	75.02
4 (student corrections)	74.12

Table 4
Results of Writing Fluency Test

Treatment Group	Adjusted Mean
1 (comments only)	125.0**
2 (corrections only)	112.1
3 (corrections and comments)	121.1
4 (student corrections)	101.1*

**highly significant (higher) p = .001
*significant (lower) p = .004

Table 5
Results of Cloze Test

Treatment Group	Adjusted Mean
1 (comments only)	34.94*
2 (corrections only)	32.91
3 (corrections with comments)	31.91
4 (student corrections)	31.91

*significant, p = .006

The results of the Writing Fluency Test showed that group 1 (comments only) was significantly more fluent, in terms of the number of words written, than the other groups. The score for group 4 (student corrections) was significantly lower than those of groups 2 and 3 (teacher corrections). (See Table 4.)

Table 5 indicates that group 1, which received only comments and no corrections, made significantly more progress in general language proficiency, as measured by the cloze test, than the groups receiving correction.

The analysis of the results of the attitude questionnaire showed a significant difference on only five items. (See Table 6.)

Table 6
Results of Attitude Questionnaire

	Item	Treatment Groups			
		1	2	3	4
	Possible score for items 1-3 is 6.0				
1.	I think that writing German gave me more confidence in speaking.	4.40	4.87	4.33	4.42
2.	I think that writing a *Tagebuch* should be retained in German 1-103.	4.81	5.13	4.42	4.27
3.	I looked forward to having my *Tagebuch* returned.	4.91	4.70	4.66	4.26
	Possible score is 2.0; 2 = 1 hr. or more, 1 = less than 1 hr.				
4.	Amount of time spent writing *Tagebuch.*	1.71	1.38	1.38	1.53
	Possible score is 4.0				
5.	Usual grade received on *Tagebuch.*	3.88	3.41	3.20	3.51

The significant differences appear as follows:

Item 1: The mean of group 2 (corrections only) is significantly higher than that of group 3 (corrections with comments).

Item 2: The combined means of groups 2 and 3 (teacher correction) is higher then 4 (student corrections).

Items 3, 4, and 5: The means of group 1 (comments only) are higher than the combined means of groups 2, 3, and 4 (corrections).

The good showing which group 2 (corrections only) made on the attitude questionnaire may be explained by the fact that a larger percentage of this group planned to continue the study of German beyond the third-quarter course, as indicated in the background questionnaire administered at the beginning of the study. (See Table 2.)

It is surprising to note that group 1 (comments only) spent significantly more time on the *Tagebuch* assignments that the other groups. One might have assumed that, since this group did not need to worry about mistakes, it would have spent *less* time rather than more on writing. This is an indication that members of group 1 still took the task seriously, and were highly motivated, even though errors did not have a negative effect on their grades.

The attitude questionnaire also asked students to make free comments regarding the writing of a *Tagebuch.* These were summarized and a tally made of the number of students of each group expressing similar opinions. (See Table 7.)

Table 7
Attitude Questionnaire Free Comments:
Summary and Tally

Comments (Summarized)	Treatment Groups 1	2	3	4
Enjoyable; gave me opportunity to write what I wanted to.	/////		/	//
Important; good idea; helpful; reinforced speaking, thinking in German, use of dictionary, and grammar.	///////	////	////	///
Should be given more weight as course component in grading.	//			
Was hard work, time-consuming.	///	//		
Length should be somewhat reduced.	///			/
There should be correction of frequently made errors; or some correction of grammar.	//////			
Should write half as many. On off week, revise and hand in again.		/		
Counting off for errors suppressed creativity and experimentation; felt pressured to stay in "safe" area.		//	//////	////
Disliked method of correction. Needed more explanation of corrections.		/	///	///////
Should be graded S/N.			/	
Difficult to think of a topic.	//		////	/
Unpleasant, tedious, not valuable.			///	//

Implications

The results of this study support the theory that correction does *not* improve students' writing skills in German as a second language, nor does it increase total competency in the language. On each of the measures of language ability, where there was significant difference among treatments, group 1, which received only comments, showed more progress than the groups which received correction.

The analysis of the attitude questionnaire revealed fewer significant differences among the treatment groups than was anticipated. Although not analyzed statistically, the free comments on the attitude questionnaire of group 1 were generally more positive toward the writing experience than those of the other groups. It is interesting to see that six students of group 1 believed that there should have been some correction of errors on their work. This coincides with the position of Kelly that, when students realize that they are communicating, and someone is understanding the message, they will *seek* advice on how to improve the mechanics of their writing.[13]

In addition to the formal evaluative instruments

used in this study, there were also unsolicited student reactions. One student of group 4 (correcting one's own mistakes), in lieu of submitting the revised version of the second *Tagebuch,* wrote a very bitter note expressing frustration at not being able to find the correct forms for the errors designated by the code symbols. The following is an exact quote: "Rather than forcing us to broaden our vocabulary as we drill our grammatical knowledge, your *'mechanische' Unterricht* will purge us of our desire to attempt any writing."

Near the end of the academic quarter a student of group 3 (corrections with comments) wrote an evaluative statement of over 200 words in which grievances regarding *Tagebuch* assignments were described. The statement began thus: "As one who very much wishes to improve their [*sic*] writing skills and basically enjoys learning the German language in general, I must express extreme disappointment and criticism of the manner in which we were supposed to 'learn' to write better." Primary among the complaints was the grading system, which discouraged attempts by students to explore new grammar and skills. This student concluded: "To continue to feel that the present program is a valuable tool in teaching students writing I feel is an insult to the intelligence of the average student. Hopefully this will be recognized, and others in the future will benefit from changes in the approach to writing."

In addition to these special notes, a number of students used the last *Tagebuch* assignment itself to express reactions to the experience of free-writing. In contrast to the special notes mentioned above, these were all positive. However, they were largely from group 1 (comments only). (See Table 8.) These students found writing a *Tagebuch* helpful, pleasant, and a good learning experience.

Table 8
Unsolicited Positive Written Comments

Group 1 (comments only)	9
Group 2 (corrections only)	0
Group 3 (corrections with comments)	1
Group 4 (student corrections)	1

There was unanimous agreement among the eight instructors that free-writing assignments had added an important dimension to the course German 1-103. Some believed that making the assignments somewhat more structured, such as at least occasionally giving assigned topics, might be an improvement.

A side benefit of treatment 1 (comments only) was that it removed the temptation for students to cheat by copying. Among groups 2, 3, and 4 there were several instances of near-certainty (if not complete certainty) that the *Tagebuch* assignments submitted were not original. This never happened in group 1.

In general, the findings of this study point to treatment 1 (comments only) as preferred and identify treatment 4 (requiring students to correct their own mistakes) as least preferred in terms of growth in language proficiency. However, this may not be due entirely to the different treatment methods *per se,* but also to the differences in the *quantity* of writing practice. In order to achieve A's, group 1 needed to write twice as much as groups 2 and 3. Group 4, on the other hand, because of revisions, wrote much less new material. This difference in quantity was most certainly one of the crucial factors in determining the difference among the groups in fluency scores. What is most unexpected is that group 4, which corrected its own errors, did not do better on accuracy, as measured by the accuracy scores on the free-writing test. This gives additional evidence that accuracy, as well as fluency, is affected positively by practice.

While the results of this research cannot be generalized beyond the group of students involved in the project, they do demonstrate that for this group of students, over a ten-week period of time, the absence of correction of free-writing assignments did not have a detrimental effect on progress in language learning, as measured by the instruments of this research. The amount of practice, on the other hand, even without correction, did appear to have a positive effect on achievement.

In summary, the initial questions posed by this research are answered as follows:

1. The time which teachers use in correcting students' original compositions is *not* well-spent. The results of this study indicate that student achievement is enhanced by writing practice alone and that corrections do not significantly increase writing skills.

2. Giving supportive comments in lieu of corrections appears to have a positive effect on student attitudes toward writing and toward the target language in general. While the answer to this question is not as unqualified as had been hoped, there is evidence to support this conclusion.

3. The addition of reinforcing comments to corrections did not make a significant difference in either achievement or attitudes as compared to correction alone. In fact, the results of the measures of both achievement and attitudes show that cor-

rection alone is equal to or superior to correction with comment. This upholds Vogler's[14] statement that a few positive comments cannot possibly counteract the negative effect of numerous corrections.

4. Students do not achieve more when they are forced to correct their own mistakes. On the contrary, the findings of this study indicate that this treatment is least effective in terms of both achievement and attitudes.

Classroom Application

The application of the findings of this research to classroom procedures would mean that the amount of writing assigned, since correction does not appear to promote competency, can be based on what is best for student learning, and need *not* be determined by the amount of time and energy a teacher has to correct it. Instead of enduring the drudgery of finding and marking errors, the teacher can, with a clear conscience, enjoy becoming better acquainted with the students through mutual sharing of information. The student, instead of being rewarded with the return of an assignment which has been mutilated by the red pen, will receive teacher responses of acceptance, encouragement, and understanding. This does not mean that corrections should never be made. When a teacher notes consistent errors, they can be explained and drilled. When students ask for correction, it can be given. Released from the burden of feeling compelled to find every error, teachers can exercise personal discretion in making corrections.

One of the most important benefits of the *Tagebuch* is the opportunity it affords the teacher to learn to know the students as persons. This is also pointed out by Staton[15] who reports on a similar approach using dialogue journals. Responding to the students' writing as in treatment 1 (with reinforcing responses and questions in the target language, so that students know that they are being understood) results in a student-teacher communication which could play a vital role in building a positive relationship. Probably more than anything else, more than any method, technique, or material used, it is this positive relationship which forms the basis for effective teaching.

NOTES

[1]Robert J. Marzano and Sandra Arthur, "Teacher Comments on Student Essays: It Doesn't Matter What You Say," a study prepared at the University of Colorado at Denver, 1977 (ERIC ED 147 864).

[2]Harriet Talmage and Maurice J. Eash, "Curriculum, Instruction, and Materials," in *Research on Teaching: Concepts, Findings and Implications,* ed. Penelope L. Peterson and Herbert J. Wahlberg (Berkeley, CA: McCutchan, 1979), pp. 161-79.

[3]Lou Kelly, "Suggestions for Helping Students Become Competent Writers," a paper presented at the Faculty Workshop on Writing, at Westmar College, in Le Mars, IA on December 11-12, 1978.

[4]Eileen N. Wagner, "When the Bookkeeping System Takes Over: The Effects of Grading Compositions on Student Attitudes," a paper presented at the Annual Meeting of the Conference on College Composition and Communication, in Philadelphia on March 25-27, 1976 (ERIC ED 120 802).

[5]Marina K. Burt, "Error Analysis in the Adult EFL Classroom," *TESOL Quarterly,* 9 (1975), 53-63.

[6]Regina T. Rinderer, "The Person in the Composing Process: A Theoretical Framework for Teaching and Research in Composition," Diss. Ohio State Univ. 1978. The information cited was gathered from the abstract.

[7]Although a total of 202 were enrolled in the course, some entered late and thus missed the pretesting. Others dropped the course during the quarter, and others were absent on the day the post-tests were given. Complete data were obtained for 70% of the potential group of students, making the subject count 141.

[8]Susanna M. Brütsch, "Convergent Discriminant Validation of Prospective Teacher Proficiency in Oral and Written Production of French by Means of the MLA Cooperative Foreign Language Proficiency Tests: French, Direct Proficiency Tests for Teachers (TOP and TWP), and Selfratings," Diss. Univ. of Minnesota 1979.

[9]Eugene J. Brière, "Quantity Before Quality in Second Language Composition," *Language Learning,* 16 (1966), 141-51.

[10]Paul W. Carstens, "Normal Cloze Scores and Multiple-Choice Cloze Scores for Testing German as a Second Language," Diss. Univ. of Minnesota 1979.

[11]John W. Oller, Jr., "Cloze Tests of Second Language Proficiency and What They Measure," *Language Learning,* 23 (1973), 105-18.

[12]Elana G. Shohamy, "Investigation of the Concurrent Validity of the Oral Interview with Cloze Procedure for Measuring Proficiency in Hebrew as a Second Language," Diss. Univ. of Minnesota 1978.

[13]See note 3.

[14]Stephen H. Vogler, "Grading Themes: A New Approach; A New Dimension," *English Journal,* 60 (1971), 70.

[15]Jana Staton, "Dialogue Journals: A New Tool for Teaching Communication," *ERIC/CLL News Bulletin,* 6, 2 (March, 1983), pp. 1-6.

Semke, H. "Effects of the Red Pen." *Foreign Language Annals* 17 (1984): 195–202. Reprinted with the permission of the American Council on the Teaching of Foreign Languages (ACTFL).