Please share your stories about how Open Access to this item benefits you.

[This document contains the author's accepted manuscript. For the publisher's version, see the link in the header of this document.]

Abstract: This study explores the effectiveness of instructor-written corrective feedback for the improvement of writing accuracy by beginning college-level learners of German. The researcher investigated changes in error rates in six error categories in essay writing in correlation with three different corrective feedback types administered consistently throughout one semester: direct, coded, and uncoded feedback. The author analyzed both short-term revision effects and semester-long changes. The study found that all groups improved their accuracy in redrafting; participants did not shorten the essay length in the final drafts to eliminate errors; direct correction led to slightly higher correction rates for selected errors; and there was no significant difference in overall error rate changes between the groups. The study concludes with suggestions for further research and pedagogical applications.

Text of paper:

The Effectiveness of Written Corrective Feedback in Teaching Beginning German

Nina Vyatkina
University of Kansas

Introduction

The debate about the usefulness of Written Corrective Feedback (WCF) in language teaching has been going on now for several decades (e.g. Ferris, 1999, 2006; Lalande, 1982; Truscott, 1996, 2007), and the results are still inconclusive. Nevertheless, teachers continue to assign students written work as a standard part of pedagogical practice, and they subsequently spend much time and effort evaluating the writing, marking and correcting errors and infelicities ranging from misspellings and missing commas to lack of organization and factual mistakes. Thus, inquiry into error correction remains of theoretical concern to foreign language researchers and of practical concern for the pedagogues. The present study will contribute to existing research by analyzing teacher feedback effects on essay writing by novice learners of German.

Please share your stories about how Open Access to this item benefits you.

Research Background

Writing accuracy has always been an important part of second language (L2)[1] teaching and learning despite continuous focus shifts in popular instructional methods. Many L2 writing researchers have argued that accuracy in writing matters to both readers and writers of L2 academic discourse (e.g., Ferris, 2006; Hedgcock & Lefkowitz, 1996; Hyland & Hyland, 2006). It is generally accepted that L2 student writers want to improve their grammatical accuracy, expect to be corrected, and value WCF from their teachers (e.g., Sheen, 2007; Truscott, 1996). Nevertheless, previous studies of WCF have yielded contradictory results. Some researchers have found evidence that systematic correction of all student errors leads to lower error rates (e.g., Lalande, 1982); others have called for selective correction of specific error types (e.g., Ferris, 2006); still others have argued that all error correction should be eliminated because it is ineffective in the long run (e.g., Krashen, 1984; Semke, 1984; Truscott, 1996, 2007). In addition, Truscott argues that the revision requirement leads students to employ the avoidance strategy and to improve accuracy at the expense of complexity.

Despite these differences of opinion, there is general agreement among researchers that WCF may be effective, but its effectiveness depends on a number of variables such as language instruction context, learners' proficiency level, and types of writing assignments. Furthermore, WCF effects may manifest themselves differently in short-term writing revisions and long-term language development. For example, Fathman & Whalley (1990) and Ferris & Roberts (2001) found positive WCF effects for short-term revisions (redrafting), whereas Robb, Ross, and Shortreed (1986) and Polio, Fleck, and Leder (1998) found no semester-long WCF effects for a similar participant population (ESL students).

Furthermore, researchers argue about the relative efficacy of different feedback types. Bitchener and Knoch (2008), providing a recent review on the subject, identify two main generally recognized WCF types: direct and indirect. Direct feedback is "the provision of the correct linguistic form or structure by the teacher to the student above or near the linguistic error" (p. 411) and indirect feedback is "that which indicates that in some way an error has been made" (p. 414). Both types can be further divided into subcategories that can be positioned on a more explicit - less explicit feedback continuum (Heift, 2010). For example, indirect feedback can be coded (marking an error with a metalinguistic code such as *WO* for "word order") or uncoded (underlining or circling of errors).

As with the question of the general effectiveness of WCF, research on feedback types has also provided conflicting evidence. Some studies found more beneficial effects for indirect WCF options (Lalande, 1982), others for direct options (Chandler, 2003), still others found no difference between the two (Robb *et al.*, 1986). Furthermore, Ferris and Roberts (2001) found no differences between coded and uncoded indirect WCF options. On the other hand, Ferris (2006) found positive long-term effects for indirect feedback given in response to "treatable" error categories. The ESL learners in Ferris's study gradually and systematically decreased the number of errors that they were able to correct on their own (e.g., in verb morphology) following indirect prompts (metalinguistic error codes) provided by the teacher. In contrast, no long-term improvement was found in "untreatable" error categories (e.g., errors made when a student attempts to use a structure not yet learned), for which teachers provided direct

Please share your stories about how Open Access to this item benefits you.

corrections (supplying correct forms).

Whereas the studies mentioned above explored the WCF effects across multiple error categories, a series of recent investigations has narrowed down the target structure to English articles (Bitchener & Knoch, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007). These carefully designed, replicable studies yielded results which converge convincingly: learners who received WCF achieved significant short-term and long-term improvement in their use of English articles, whereas the control groups did not. Notably, WCF was equally effective for migrant and international ESL students (Bitchener & Knoch, 2008) as well as EFL students (Ellis *et al*., 2008).

The latter studies are representative of an overwhelming majority of existing WCF research that has concentrated on intermediate to advanced L2 English students enrolled in university composition courses. In contrast, WCF research on languages other than English remains extremely scarce. For example, Kepner (1991) explored writing accuracy of two groups of intermediate learners of Spanish: the experimental group received direct corrections and the control group received only content-related feedback. The researcher found no differences in accuracy changes over the course of a semester. Mantello (1997) compared the effects of indirect error corrections (coded WCF) and reformulations (rewriting of learner essays by a native speaker) for Canadian 8th graders who learned French as an L2 in an immersion environment. Both groups improved their ability to identify and produce the target grammar structure (the narrative past tense) over four months, and neither group outperformed the other on the tests.

For teaching German as a foreign language (GFL), two early studies have compared WCF methods and their effects on writing of intermediate learners of German at American universities: Lalande (1982) and Semke (1984). Lalande (1982) compared the direct and the indirect WCF options. For the direct group, the teacher wrote all corrections onto student compositions. For the indirect group, the teacher marked the errors by means of an error correction code. Findings showed that the indirect group outperformed the direct group in most of the non-lexical error categories. Semke (1984) compared the effects of four different methods of teacher treatment: 1) writing comments and questions rather than corrections; 2) marking all errors and supplying the correct forms; 3) combining positive comments and corrections; 4) indicating errors by means of a code. Results indicated that all groups improved their accuracy over the course of a semester but no condition was more beneficial than the others. In sum, Lalande (1982) claimed greater benefit from more explicit feedback, whereas Semke's (1984) study showed no difference between more or less explicit feedback.

Recent advances in computer-assisted language learning (CALL) have brought about new forms of WCF, ranging from teachers using computers for providing feedback to students making revisions to automatic WCF provided by intelligent language tutoring systems (ILTS). Emerging WCF research in CALL environments has the potential to shed more light onto controversial issues in the field (see Hyland & Hyland, 2006, for a review). Comparing studies of paper-and-pencil and computer-based L2 writing revisions, Arnold, Ducate, and Kost (2009) conclude that learners make more revisions on computers because, even for less proficient language learners, it is easier to correct, add, and delete parts of

Please share your stories about how Open Access to this item benefits you.

---

electronic texts. The authors explored the behavior of intermediate learners of German who participated in peer-editing in a collaborative computer-based environment (*Wiki*). The results showed that 76% of all attempted revisions were successful and that the group that received teacher feedback in addition to peer feedback achieved higher rates of accuracy changes and had more successful revisions.

A number of recent CALL studies have continued research on relative effectiveness of WCF options. Sauro (2009) explored the impact of metalinguistic feedback (explanations about the nature of the error) and reformulations via text-chat on the use of English articles by intermediate and advanced learners who were native speakers of Swedish and found no advantage for either feedback type, although the metalinguistic group performed significantly better than the control group on the immediate posttest. Heift's multiple studies (e.g., 2008, 2010) investigated various aspects of computer-assisted WCF in teaching GFL. These contributions are especially valuable given the paucity of studies on languages other than English. Among other questions, Heift (2010) researched the effects of two different types of automatic WCF provided by the *E-Tutor* ILTS on learner immediate self-correction of grammar and spelling errors, or uptake. The feedback types under investigation were metalinguistic explanations (ME) and metalinguistic clues (MC). The ME feedback informs the students about the specific type of error they made (e.g.: "you made a mistake with this article. The noun is neuter"), whereas the MC feedback only gives the student a hint as to the general error category (e.g.: "Grammar?" or "Spelling?"). The results show that for beginning learners, there was no difference in self-correction rates in response to the error-specific feedback type and to the generic feedback type. However, over the course of three semesters, the learners considerably increased their error-correction rates in response to the more explicit WCF (ME), whereas their uptake did not change much over time in response to the less explicit WCF (MC). The findings indicate an advantage for more explicit WCF as students gradually become more familiar with metalinguistic terminology and with the CALL program.

Given the inconclusiveness of WCF research reviewed above, one may wonder what decisions L2 practitioners make in handling student writing. O'Donnell's (2007) study reports the results of a survey about current policies and practices in foreign language writing administered to American university language program directors. More than two thirds of the respondents wrote that they required multiple drafts for graded compositions (p. 664). However, no uniform WCF policies existed and/or were being followed at most institutions. Respondents were also unanimous in their statement that the preferred WCF type in their FL programs was coded feedback. On the other hand, 75% of the participants stated that their programs did not have a uniform policy about WCF format that instructors were expected to give to students (p. 661). In general, the study showed that language program directors as well as instructors in many programs invest much effort into commenting on multiple drafts of student writing, but many of them are uncertain as to what WCF types are more beneficial to students at what proficiency levels. O'Donnell (2007) concludes her study by a call for a dialogue between FL writing researchers and educators in order to work out a best practice policy in the field.

To summarize, although the question of the general WCF effectiveness and the relative advantages of different WCF options remains to a large extent unresolved, there are some indications that 1) teacher feedback is valued by students and helps them to achieve better accuracy on revisions; 2) WCF may be

4

Please share your stories about how Open Access to this item benefits you.

effective for selected error categories, especially when it is focused; 3) as long as explanations are clear to students, explicit and more specific metalinguistic WCF is more beneficial; 4) regularly provided computer-assisted feedback may be more conducive to accuracy improvement. Without exception, all studies reviewed above call for future (especially longitudinal) research targeting a variety of grammar structures, proficiency levels, and languages. Responding to these calls, this study explores the relative effectiveness of several WCF types in first-semester L2 German and addresses existing research gaps by expanding the empirical research basis in languages other than English (Ellis *et al.*, 2008), comparing feedback impact on student progress in relation to different error types (Ferris, 2006), and targeting beginning learners (Elola, 2008).

Design of the study

Research objectives and research questions

The main objective of this study was to investigate the impact of different WCF types on the improvement of writing accuracy of beginning GFL learners. More specifically, the focus of the study was on error rate changes at several points in time during one academic semester. Short-term error rate changes were measured between the rough draft and the final draft of three essays for the cumulative error rates and for specific error categories. Long-term error rate changes were explored in different pieces of writing (between the rough drafts of three essays). It was not expected that learners in their first-semester language course would reduce the overall number of errors in the last essay written at the end of the semester in comparison to the first essay. In contrast, it was expected that students would practice newly learned L2 structures in each new piece of writing and, therefore, make more errors. It was also expected that specific writing tasks would induce errors in certain categories. Therefore, this study did not focus on absolute error reduction or increase rates. Instead, it investigated whether the variation in frequencies of different error categories between the three time points would correspond differently to three different feedback types. In addition, a comparison of the length of each essay draft pair was carried out before performing comparisons of error frequencies to explore whether learners employed a strategy of avoiding error correction by simply cutting the passages marked by their instructors as requiring a revision.

Thus, the following research questions were formulated:

1. Do revised essay drafts differ in length from rough drafts?
2. Do error rates change differently in response to different WCF types in redrafting?
3. Are different error categories affected differently by error treatment in redrafting?
4. Do error rates change differently in response to different WCF types over the course of a semester?

Participants

The participants in the study were 66 students at a large public Midwestern university enrolled in the first-semester German and their instructors, five graduate teaching assistants. Most students were

5

Please share your stories about how Open Access to this item benefits you.

undergraduates, and one third of all participants (22 people) were first-semester freshmen. The age of 56 participants ranged from 18 to 23 years with an average of 19.5 years, and 10 participants were 25 or older. 24 students (36%) were female, and 42 (64%) were male. Two thirds of the participants stated in the pre-course questionnaire that they had not studied German before, whereas one third (22 participants) indicated that they had had 0.5 – 2.5 years of German in High School. These 'false beginners' either took a placement test before the course and scored too low to be placed in a higher level course or decided on their own that they needed a fresh start with German. All participants wrote that English was their first language (L1) with the exception of one student whose L1 was Turkish. Since there was variability between subjects in previous exposure to the target language as well as their first language, a pretest was administered to check whether the groups had comparable error frequencies at the beginning of the study. No significant differences in the pretest scores were found (see the "Results" section).

The teaching assistants were three female native speakers of German and two male native speakers of English. All instructors were first or second semester graduate students of German (four of them at the M.A. level and one at the Ph.D. level) at the same university. Four instructors had some experience in teaching GFL and one instructor (a native speaker of German) had previously taught EFL.

Pedagogical experiment

During fall semester 2008, under the researcher's supervision, five graduate teaching assistants, who taught six sections of first-semester college German, consistently used one of three types of feedback in correcting the five paragraph-long essays students wrote during the semester. The feedback types employed in the experiment were: (a) *direct correction*, underlining the errors and providing the correct form; (b) *coded feedback*, indicating the type of error based on an abbreviated code system; (c) *uncoded feedback*, underlining the errors without specifying their type (Robb *et al.*, 1986). Sections were randomly assigned to the feedback types. Each two sections out of six received one of the three WCF treatments. Four instructors taught one experimental section each, and one instructor taught one coded WCF section and one uncoded WCF section.

The course met five days a week for 50 minutes during a 16-week-long semester, with one hour spent weekly in a computer lab. In-class time was mostly devoted to task-based interactive activities in form of pair work and group work, but also included grammar explanations and practice (see, e.g., Brown, 2007). Computer lab time was used for web-based activities and essay writing. During the semester, students completed five two-draft compositions on topics provided at the end of each chapter of the course workbook (Briggs, Di Donato, Clyde, & Vansant, 2008) and based on the material covered in the corresponding textbook chapter (Di Donato, Clyde, & Vansant, 2008). The students typed the first draft in the computer lab, saved it electronically, and submitted it through the online courseware (*Blackboard*). In this way, they were writing under controlled timed conditions, under the instructor's supervision, and were allowed to use online dictionaries but not online translators. The students were required to write approximately 70 words for each essay. The instructors retrieved the essays from *Blackboard*, provided electronic WCF according to the assigned treatment condition, and returned

6

marked essay drafts to the students via *Blackboard*. Students revised the essays outside of class and submitted the final drafts electronically. Then, instructors gave electronic WCF on the final drafts keeping the treatment condition constant and returned them to the students. In addition, students received grading sheets for each essay draft filled out by instructors. Half of the grade points were allocated by a holistic evaluation of content, relevance, creativity, and complexity, and the other half was assigned for grammatical accuracy, word choice, and spelling. The first draft was worth 70% and the revised final draft 30% of the total essay grade.

All WCF in this study was provided electronically. For the uncoded WCF, words and phrases containing errors were underlined using the underline feature in Microsoft Word. For direct WCF, words and phrases containing errors were also underlined, and a suggested accurate form was typed by the instructor in parentheses following the erroneous form. For the coded WCF condition, an electronic toolbar developed for the purposes of this study allowed the instructor to underline the erroneous form and to insert a colored error code as a superscript right after it. The instructors tagged errors in accordance with an error coding chart (see Appendix) that included codes for 15 specific error types belonging to the categories commonly used in language teaching practice and research: lexical choice, noun-related errors, verb-related errors, spelling, and structural errors (e.g., Arnold *et al.*, 2009; Ferris, 2006). Each category included several specific error types. For example, the errors "ending", "gender", and "case" comprised the category of noun-related errors. The codes for the error categories were marked in different colors: for example, all noun-related errors were assigned red codes, and all verb-related errors were assigned blue codes. It was assumed that this visual enhancement would help beginning language learners in perceiving and understanding metalinguistic codes. The coding chart with explanations was handed out to all students in the coded treatment group, posted in *Blackboard*, and repeatedly explained by instructors in class.

Data coding

The five instructors teaching the experimental classes were also research assistants in this study. This activity constituted an action research component of the graduate course in Teaching Methodology, in which they were enrolled during the same semester.

Consistency of the coding procedure between instructors was insured by discussions and joint coding sessions during weekly meetings of the methodology course and by spot-checks by the researcher. After having returned essay drafts to students marked according to one of the three WCF conditions, all instructors coded electronic copies of student drafts for 15 error types using the electronic toolbar and submitted coded drafts to the researcher. In other words, students in each experimental section received their essays from instructors with errors either underlined (uncoded condition) or coded (coded condition), or followed by suggested corrections (direct condition), whereas the researcher received essays with all errors coded. In other words, teachers planted electronic error tags into student essays that served as "clues or bootstraps" (Wible, Kuo, Chien, Liu, & Tsao, 2001) for subsequent learner language analysis by the researcher. Only rough and final drafts of essays 1, 3, and 5 were coded by the

instructors for this study, thus representing student writing at the beginning, in the middle, and at the end of the semester.

It has to be noted that student revisions were not coded "as successful or unsuccessful to indicate whether the change resulted in an error-free form with regard to the aspect that the revision addressed" (Arnold *et al.*, 2009, p. 125). Thus, this study did not seek to explore how many particular errors were corrected by the students in their final drafts in comparison with rough drafts. Instead, the focus of investigation was on overall error rate changes from draft to draft and in new pieces of writing.

Data pool

All coded essays have been archived in an electronic database. The total amount of data coded by the teachers and subjected to the analysis consisted of 324 texts (rough and final drafts of three essays) written by 66 students and containing a total of 28078 words. The mean number of words in a text was 86.66 words, with a range of 46 to 247. In these texts, the teachers marked the total of 2385 errors and provided learners with written corrective feedback on each error according to the treatment condition assigned to them. Each text contained on average 7.36 errors. The distribution of errors according to error categories was as follows (in the descending order): 704 lexical errors (29.52%), 680 noun-related errors (28.51%), 383 spelling errors (16.06%), 240 verb-related errors (10.06%), 166 word order errors (6.96%), 125 sentence structure errors (5.24%), and 87 punctuation errors (3.65%).

Results

RQ1: Do revised essay drafts differ in length from rough drafts?

The database software used for data collection in this study (*FilemakerPro*) performed an automatic word count of archived texts. To answer RQ1, the length of each student text was measured in number of words, and repeated measures ANOVAs were run for these frequencies.

The descriptive statistics for the mean essay length are presented in Table 1. The ANOVAs revealed no significant differences between the length of the two drafts of each of the three essays: $F(1, 58) = .53$, $p > .05$ (.819) (essay 1); $F(1, 37) = .69$, $p > .05$ (.412) (essay 3); $F(1, 31) = 1.871$, $p > .05$ (.181) (essay 5).

	Group	coded condition		uncoded condition		direct condition	
Draft	**N**	**M**	**SD**	**M**	**SD**	**M**	**SD**
E 1.1	61	78.18	14.27	73.47	11.37	76.33	14.42
E 1.2	61	78.82	13.89	72.33	11.13	77.12	14.84
E 3.1	40	104.12	42.73	103.73	32.76	112.00	22.01
E 3.2	40	104.06	42.32	102.55	33.22	112.08	20.49
E 5.1	34	81.19	20.86	81.56	18.35	77.33	12.51
E 5.2	34	80.88	20.49	87.56	15.72	75.78	14.97

Table 1 Descriptive statistics for the essay length (number of words)

Furthermore, the ANOVAs revealed no interaction effect between time and treatment for essays 1 and 3: $F(2, 58) = 1.835$, $p > .05$ (.169) (essay 1) and $F(2, 37) = .657$, $p > .05$ (.524) (essay 3). There was an interaction effect for essay 5: $F(2, 31) = 4.856$, $p < .05$ (.015), $\eta^2 = .239$, which may be explained by the fact that the uncoded group's final drafts were on average longer than the final drafts of the other two groups. However, this difference was negligible because the post hoc Bonferroni comparison did not show a significant effect for any pairs of treatment.

RQ 2: Do error rates change differently in response to different WCF types in redrafting?

Before the analysis, error frequencies were normalized per 100 words (see Fig. 1).

$$relative\ error\ frequency = \frac{raw\ error\ count}{total\ word\ count} \times 100$$

Figure 1 Formula for calculating the relative error frequency

The relative (normalized) error frequencies on the rough drafts of the first essay were used as the pretest scores. One-way ANOVAs were run for all seven error categories. The results showed no statistically significant differences in the pretest scores (sum of errors in each category per 100 words) among the three groups except for the "punctuation" category. For this reason, this category was excluded from all subsequent analyses and only six error categories were considered: verb-related, noun-related, lexical, structural, word order, and spelling errors.

The descriptive statistics for the mean normalized overall error scores (sum of all errors per 100 words) are presented in Table 2 and illustrated in Figures 2, 3, and 4.

	Group	coded condition		uncoded condition		direct condition	
Draft	**N**	**M**	**SD**	**M**	**SD**	**M**	**SD**
E 1.1	61	12.22	6.32	18.65	12.99	18.90	12.58
E 1.2	61	3.50	2.49	8.24	7.99	2.47	4.33
E 3.1	40	11.22	9.55	9.78	6.67	11.90	10.77
E 3.2	40	4.44	7.91	4.10	5.15	1.54	1.63
E 5.1	34	16.54	12.37	17.06	7.63	23.84	7.87
E 5.2	34	3.95	5.53	6.89	3.50	3.77	5.55

Table 2 Descriptive statistics for relative frequencies for the sum of all errors

The ANOVAs revealed a significant short-term change from draft to draft with a considerable effect size for all three essays:

E1: $F(1,58) = 8.777$; $p < 0.05$ (.000); $\eta^2 = .602$

E3: $F(1,37) = 4.472$; $p < 0.05$ (.000); $\eta^2 = .547$

E5: $F(1,31) = 7.993$; $p < 0.05$ (.000); $\eta^2 = .721$

The graphs in Fig. 2, 3, and 4 demonstrate that the revised drafts of all three essays for all treatment groups had significantly fewer errors than the rough drafts. In addition to these results, the ANOVAs revealed no interaction effect between time and treatment for essays 3 and 5: $F_{(2,37)} = 1.487$; $p > 0.05$ (.239) (essay 3) and $F_{(2,31)} = 3.077$; $p > 0.05$ (.06) (essay 5). There was an interaction between time and treatment with a small effect size for essay 1: $F_{(2,58)} = 3.974$; $p < 0.05$ (.024); $\eta^2 = .121$. However, the post hoc Bonferroni comparison did not show a significant effect for any pairs of treatment.
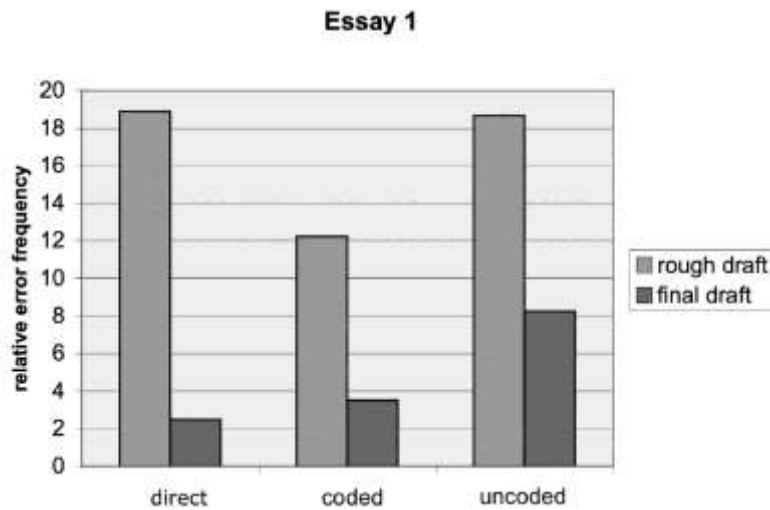

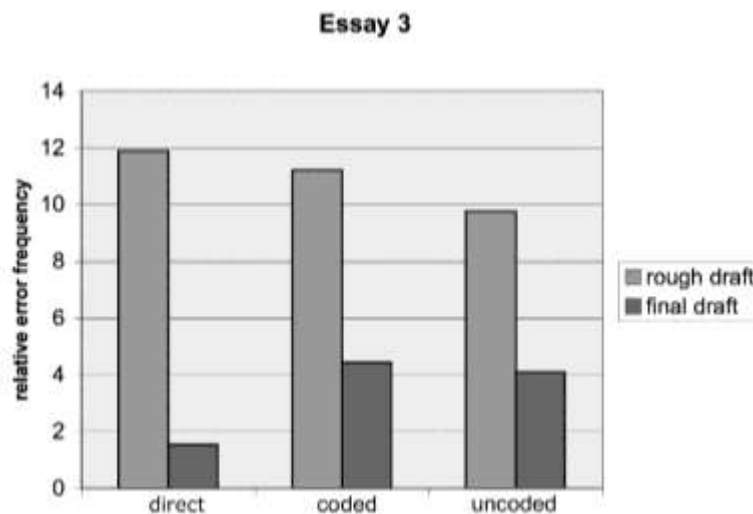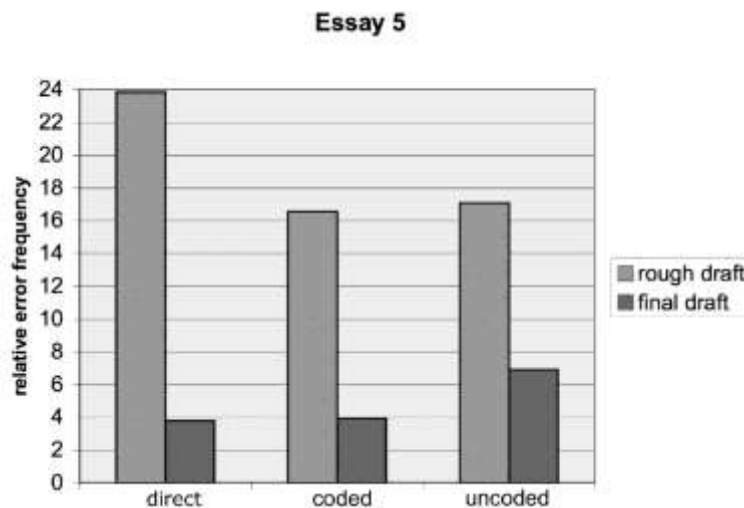**Figure 2** Relative error frequencies in the rough draft and the final draft of essay 1

Figure 3 Relative error frequencies in the rough draft and the final draft of essay 3

11

Figure 4 Relative error frequencies in the rough draft and the final draft of essay 5

RQ 3: Are different error categories affected differently by error treatment in redrafting?

The time/treatment comparisons were also performed for each of the specific error categories for each pair of drafts. The results showed that the accuracy significantly (p < .05) changed for each of the six focal error categories from draft to draft of all three essays (in the 17 comparisons out of 18). The only exception was essay 3 in which there was no significant change for word order. More specifically, the error rate of each error type decreased from draft to draft in most cases (which was expected).

Furthermore, the ANOVAs revealed no interaction between treatment and time for most comparisons. In other words, all three feedback types led to similar error rate changes between drafts in each specific error category. There were three exceptions to this pattern that showed a treatment/time interaction with small to medium effect sizes: for the word order in essay 1: $F(2,58) = 4.954$; $p< 0.05$ (.01), $\eta^2=.146$; 2) for spelling in essay 3: $F(2,37) = 3.890$; $p< 0.05$ (.029), $\eta^2=.174$; and 3) for structural errors in essay 5: $F(2,31) = 7.094$; $p< 0.05$ (.003), $\eta^2=.314$. These interactions were explored further by post-hoc pairwise comparisons. In essay 1, the interaction effect occurred because the word order accuracy increase was significant only for the direct and coded group but not for the uncoded group. In essay 3, the spelling accuracy significantly increased for the direct and uncoded group but only approached significance (p= .06) for the coded group. Finally, in essay 5, only the direct group performed significantly better in the structural accuracy on the second draft.

RQ 4: Do error rates change differently in response to different WCF types over the course of a semester?

To answer this research question, the time/treatment comparisons were performed for each of the specific error categories. Frequencies were measured for the rough drafts of essays 1, 3, and 5 that were completed by 56 subjects. No effects were revealed for the categories Verb, New Structure, and

12

Spelling, i.e. the error rate changed to a comparable amount for all groups over the course of the semester. There was a time effect with a moderate effect size for the categories Noun: $F(2,52) = 2.575$; $p < 0.05$ (.000), $\eta^2 = .498$ and Word Order: $F(2,52) = 1.393$; $p < 0.05$ (.000), $\eta^2 = .349$. However, there was no interaction between time and treatment. More specifically, learners in all groups made more noun-related errors at the end of the semester in comparison to the beginning of the semester (Fig. 5) and fewer word order errors on essay 3 than on essays 1 and 5 (Fig. 6). Finally, for the Word category, there was a time effect: $F(2,52) = 7.356$; $p < 0.05$ (.002), $\eta^2 = .221$ and there was an interaction between time and treatment: $F(4,106) = 3.506$; $p < 0.05$ (.01), $\eta^2 = .117$. The effect sizes for both results were small, and the pattern of error rate change was mixed: most variation appears to fall on essay 1, whereas all groups made a similar amount of errors on essays 3 and 5. Therefore, no apparent differences between any of the treatment conditions were revealed.

Discussion

The purpose of this study was to investigate whether beginning learners of German improved their writing accuracy in the result of the revision and whether one WCF type had more short-term and long-term advantages than the other two.

The first research question explored whether participants in this study reduced the essay length during the redrafting process instead of correcting errors marked by the teacher. The analysis showed that all draft pairs in all treatment groups were not significantly different in length. In other words, students performed revisions without either shortening or lengthening their essays in the final draft, and did so irrespective of the WCF condition. This finding contrasts with Truscott's (2007) claim that "corrected students tend to shorten and simplify their writing […], apparently to avoid situations in which they might make errors" (p. 268). This result is especially interesting since the learners in this study had some room for using this strategy: all groups wrote more than the required 70 words on each rough draft (especially for essay 3 that averaged about 105 words), and thus could have cut some erroneous structures instead of correcting them without risking a grade reduction. In other words, the participants in this study did not improve their accuracy at the expense of the complexity (as measured in length).

This study also showed that the revision process helped students to improve the accuracy in redrafting. Instructors in this study found 55% - 88% fewer errors in the final drafts than in corresponding rough drafts[2]. Moreover, separate tests performed for each error category yielded similar results. Participants in this study made significantly fewer errors related to verbs, nouns, word order, and spelling as well as lexical and structural errors in the revised drafts. This finding corroborates previous research results showing that self-editing leads to superior final writing products (e.g., Polio *et al.*, 1998; Ferris, 2006).

Next, individual error categories were tested separately for WCF effects. In particular, it was expected that lexical and structural errors might not be amenable to student self-correction in response to indirect feedback types. However, although the direct condition was found to be more beneficial for improving the word order, spelling, and structural organization on separate essays, all three feedback types led to similar accuracy improvements between drafts in each specific error category in most cases.

This finding supports those from previous research showing similar error reduction rates for various WCF conditions (e.g., Bitchener & Knoch, 2008; Ellis *et al.*, 2008, Ferris, 2006; Sheen, 2007).

Finally, error rates in new pieces of writing (rough drafts of essays 1, 3, and 5) in response to different WCF types were compared. Similarly to short-term changes, no significant long-term differences were found between treatment conditions in either error category. Frequencies of verb-related, structural, and spelling errors did not change significantly over the course of the semester in any group. Noun and word order error frequencies changed similarly in all groups. All learners made more noun-related errors at the end of the semester, which may be explained by learning and practicing increasingly more complex noun morphology, and fewer word order errors on essay 3, which may be attributed to a task effect (essay topic). Lexical errors showed a mixed pattern with much variation on essay 1 and no significant group differences later on. To illustrate the effects of different feedback types, some concrete examples of how individual learners responded to teacher markings will be considered below (Fig. 5, 6, 7).

Not surprisingly, the direct feedback (explicit correction of learner errors) led to the highest accuracy improvement rates between drafts. All the students had to do was to delete their error highlighted by the teacher, delete the parentheses and thus incorporate the teacher's suggestion into the final essay draft. However, as seen in Fig. 5, this technique may lead to mechanical changes performed by students. In 1.1, the student did not remove the erroneous form from the rough draft (*ist*) and used it along with the teacher's suggestion (*sind*). In 1.2, the student pasted the suggested correction into a wrong place, which resulted in two incomprehensible sentences. Obviously, if students perform surface changes without thinking and understanding what they are doing, such editing does not contribute to their learning and only leads to teachers' wasted time and effort.

Rough draft with teacher markings ( )	Final draft - student revision
1.1. Mein Bruder und meine Schwester sind Zwillinge und sie ist (**sind**) 18 Jahre alt. 'My brother and my sister are twins and they is (are) 18 years old.'	Mein Bruder und meine Schwester sind Zwillinge und sie ist sind 18 Jahre alt. 'My brother and my sister are twins and they is are 18 years old.'
1.2. Ich heiße Barb (**Ich bin im**) Mai 1980 geboren. 'My name is Barb. (I was) born (in) May 1980.'	Ich bin im heiße Barb. Mai 1980 geboren. 'I was in name is Barb. May 1980 born.'

Figure 5 Examples of direct feedback

14

Please share your stories about how Open Access to this item benefits you.

Rough draft with teacher markings < >	Final draft - student revision
2.1. Er ist ein </W> Informatiker von Beruf. *'He is a computer scientist.'*	Er ist einen Informatiker von Beruf. *'He is a [ending] computer scientist.'*
2.2. Ich habe […] ein paar </Sp> Socke </E> und Schuhe. Die Hose Farbe </Sp> sind</VF> „Black Moss". Die Socken ist </VF> schwarz oder weiß. Ich brauche […] braun </E> oder schwarz </E> Schuhe! Ich mag mein Lieblingskleidungsstück, und ich tragen </VF> es meistens am Sonntag zu </E> Kirche. *'I have […] a couple [spelling] of socks [ending] and shoes. The color of the pants [spelling] are "Black Moss". The socks is black or white. I need […] brown [ending] or black [ending] shoes! I like my favorite piece of clothes, and I wear [plural] it mostly on Sunday to [ending] church.'*	Ich habe […] ein päar Sock und Schuhe. Die Hose Farben ist „Black Moss". Die Socken sind schwarz oder weiß. Ich brauche […] braune oder schwarze Schuhe! Ich mag mein Lieblingskleidungsstück, und ich trage es meistens am Sonntag zur Kirche. *'I have […] a couple [spelling] of sock [ending] and shoes. The color of the pants [spelling] is "Black Moss". The socks are black or white. I need […] brown or black shoes! I like my favorite piece of clothes, and I wear it mostly on Sunday to church.'*

Figure 6 Examples of coded feedback

Rough draft with teacher markings	Final draft - student revision
3.1. Ich bin am neunzehn. Mein Hobbys sind Computerspiele spielen und Karten spielen. Für Spaß ich mache es Reisen und es Wandern. *'I am at nineteen. My [singular] hobbies are to play computer games and to play cards. For fun, I do it traveling and it hiking.'*	Ich bin neunzehn. Ich Hobbys sind Computerspiele spielen und Karten spielen. Für Spaß ich wie Reisen und Wandern. *'I am nineteen. I hobbies are to play computer games and to play cards. For fun, I as traveling and hiking.'*
3.2. Ich bin geboren August 20, 1987. Ich bin 1,83 meter groß. Ich Computerspiele spiele. *'I was born [word order] August 20, 1987. I am 1,83 meters [spelling] tall. I computer games play.'*	Ich bin gebore August 20, 1987. Ich bin 1,83 Meter groß. Ich Computerspiele spielen. *'I was born [word order, ending] August 20, 1987. I am 1,83 meters tall. I computer games play [plural].'*

Figure 7 Examples of uncoded feedback

The coded feedback indicated the location of errors and provided students with metalinguistic clues for correction. However, sometimes it led to students' wrong guesses and erroneous corrections, as shown in Fig. 6. In 2.1, the code "W" indicated that the article *ein* was superfluous in the given context. However, the student misinterpreted the clue and added an ending to the article (*einEN*) instead of deleting the word. In 2.2, the "Sp" (spelling) codes were not sufficient for the student to perform successful corrections. Instead of capitalizing the "p" in *paar*, he replaced an "a" with an a-umlaut to arrive at *päar* instead of "Paar". Similarly, he replaced *Hose Farbe* with *Hose Farben* instead of the accurate *Hosenfarbe*. It can be assumed that underlining specific erroneous letters or word parts instead of whole words might have given this student more easily interpretable clues and thus led him to more successful corrections. On the other hand, this student successfully utilized the VF codes hinting at erroneous verb forms and correctly changed the plural verb endings to singular endings and vice versa

15

Please share your stories about how Open Access to this item benefits you.

(2.2). Finally, the E codes indicating erroneous noun, adjective, and article endings, led this student to accurately add three endings (*braunE, schwarzE, zuR*) but to inaccurately delete one ending instead of changing it (*SockE -> Sock* instead of *SockEN*).

The uncoded feedback, as the most implicit WCF type, only indicates the location of errors to the student who has to decide what exactly needs to be corrected. Fig. 7 shows that the student in 3.1 correctly deleted the superfluous underlined word *am*. However, this student was not able to interpret the underlined *Mein* as a hint for adding an ending and changed the possessive pronoun to the personal pronoun *Ich* instead. Similarly, the underlines were not sufficient to make it clear that the verb *mache* needed to be moved to the 2nd position in the sentence and the two pronouns *es* needed to be deleted. Instead, the student deleted all underlined words and replaced them with the word *wie*, thus arriving at an incomprehensible sentence. In 3.2, the implicit feedback led the student to successfully correct the spelling of *meter* (capitalization of the first letter) but to unsuccessfully change endings of the verb forms *geboren* and *spiele* instead of changing the word order.

These few considered examples give a glimpse into the complexity of the WCF issue. Even when teachers do their best in providing consistent feedback and students put a lot of effort into interpreting and incorporating teacher markings, the revision results may be often unpredictable, leading to successful changes in some cases and to wrong guesses and therefore unsuccessful changes in others.

Limitations

The statistical results of this study should be interpreted with caution.

First, the standard deviations in this study were quite large (Table 2) as is typical for this sort of research (e.g., Ferris, 2006). It is worth noting that the standard deviations for the beginning learners in this study did not reach the magnitude found for more proficient learners in previous research[3]. However, that the variance was still quite high suggests the necessity of using additional methods[4] for exploring learner development that address individual variation and variation between proficiency levels (cf. Heift, 2008). Case studies using qualitative methods could shed more light onto microchanges occurring in individual students' writing over time (Ortega & Byrnes, 2008). However, the high level of variance found in this study should be considered as a result in itself, showing that none of the feedback methods triggered consistent improvement of writing accuracy by all students.

Second, this study focused on error rates as marked by teachers in student writing. Future studies using semi-automatic error annotation (Heift & Rimrott, 2009) should be performed on learner corpora (large electronic collections of learner texts) to confirm or refute findings from this study and other existing WCF research to arrive at more generalizable results. Subsequent analyses should compare corrections of particular errors in redrafting (cf. Fig. 5, 6, 7) in addition to analyses of overall accuracy improvement rates.[5]

Finally, this study did not use a control group because of the ethical concern of requesting a second draft from the students if no feedback was given, so no claims can be made as to whether WCF is more

Please share your stories about how Open Access to this item benefits you.

effective than no feedback for successful revisions and long-term writing improvement.

Pedagogical implications

One of the inspirations for this study was the need of L2 educators to decide how to administer effective and efficient WCF to student writing. This need is especially urgent in beginning L2 courses since WCF research targeting low proficiency levels is virtually non-existent and these courses are frequently taught by novice teachers (graduate assistants) who require much guidance. This study found that first-semester students are able to significantly improve the accuracy of their writing in the redrafting process in response to teacher feedback without cutting the length of their essays. This finding speaks in support of the process approach to writing involving multiple drafts even at beginning levels. However, no significant advantages were found for any particular feedback type. Therefore, Ferris's (2006) finding that mere indication and location of error was sufficient for intermediate ESL students for successful revisions appears to be supported also for beginning FL students. On the other hand, there were some indications that direct feedback led to more successful revisions, at least for separate error categories in some essays. Concrete examples of revisions performed by individual students showed that more implicit feedback types (coded and uncoded) may be confusing for learners and prompt them to make wrong guesses, which confirmed similar findings from previous research (e.g., Hedgcock & Lefkowitz, 1996).

Given the fact that no major advantages were found for any feedback type, what recommendations can be given to L2 practitioners? Does it mean that feedback to beginning learners should be administered randomly or following the intuition of each particular instructor? The main implication of this study is that no feedback condition is beneficial if applied alone without strategically combining it with others. Admittedly, direct correction leads to better immediate improvement rates but it is questionable whether it is more beneficial for long-term acquisition. Therefore, language educators should not choose the easy way and decide in favor of using exclusively one WCF type (e.g., coding) across the board. Instead, the instructors should be trained in separating the errors that students are supposed to correct on their own from "untreatable" errors at each particular stage of the instruction process. As Bitchener and Knoch (2008) and Ferris (2006) suggest, direct feedback should be reserved for "untreatable" errors, e.g., words and structures that learners are not familiar with. However, teacher suggestions in direct corrections should be kept to a minimum in order to avoid the appropriation of student texts (Hyland & Hyland, 2006). Furthermore, most "treatable" errors should be marked by easily interpretable codes. For example, the very beginners in this study were able to correct the subject-verb agreement errors in response to the coded feedback. Similarly, word order errors should be consistently marked with a code on the verb because the uncoded feedback was shown to lead to deletions or replacements of the underlined word instead of position changes. On the other hand, spelling errors and adjective ending errors may be amenable to student editing if teachers simply underline erroneous word parts. Moreover, as many WCF researchers (e.g., Ferris, 2006; Lee, 2008; Mantello, 1997; Melin, 1998) suggest, teachers should not be required to mark each and every error in student writing. Teachers who participated in this study commented in informal discussions that marking all errors following one and only one method was tiring and counterintuitive. These comments in conjunction with this study's

results suggest that feedback should rather concentrate on several focal error types in accordance with the syllabus and/or on errors constituting problem areas for particular students.

Another pedagogical implication arising from the results of this study and other existing research relates to the need for integration of teacher WCF with other methods of enhancing learner accuracy. It was apparent from the examples considered in this study that no feedback type alone was unambiguous enough for learners to make consistently successful revisions of all errors. In light of this finding and considering how time-consuming giving feedback is for teachers, it may be recommended to diversify writing assignments and to alternate those requiring teacher WCF with other types. The following specific changes have been implemented in the researcher's own language program based on this study's results which have been positively evaluated by both students and teachers and may serve as examples for other teachers. First, peer review and peer-editing sessions have been introduced in addition to teacher WCF (see also Arnold *et al.*, 2009; High, Hoyer, & Wakefield, 2002, 2002; Melin, 1998). Second, fewer but longer essays involving multiple drafts and teacher WCF have been supplemented with more frequent shorter journal writing tasks with no error correction but short text-specific marginal and end comments (both content and accuracy related).[6] Finally, the synergy between feedback provided by teachers and by the intelligent language tutoring system (ILTS) *E-Tutor* (Heift, 2008, 2010) implemented in the program seems especially promising. The *E-Tutor* is a tireless WCF-giver that leads the student through stages of feedback given on each particular error while automatically adjusting the level of directness-indirectness according to the performance level. Moreover, learners of German who consistently receive WCF from the *E-Tutor* in sentence-building and translation practice tasks have been shown by Heift and Rimrott (2009) to improve their word order accuracy rate in the essays written upon completion of each practice task segment. Teachers are advised to supplement their teaching with ILTS because it allows systematic correction of surface-level learner errors to be shifted to individualized student-computer interaction, while valuable teacher effort and time can be devoted to higher-level feedback provided on content, genre, textual organization, and appropriateness of student texts (see also Byrnes, 2009; Elola, 2008).

Acknowledgements

Please share your stories about how Open Access to this item benefits you.

Notes

1. L2 is used henceforth as a generic acronym to refer to second language, irrespective of whether this language is an official or major language in the country of instruction or not. If this difference needs to be underscored, the acronyms such as ESL (English as a Second Language), EFL (English as a Foreign Language), or GFL (German as a Foreign Language) are used.

2. There were indications that the direct condition led students to immediately improve their accuracy rate to a greater extent (ca. 85%) than the coded condition (ca. 70%) and the uncoded condition (ca. 60%). This result mirrors the findings by Ferris (2006) for intermediate ESL students. However, no statistically significant differences in error rate changes in redrafting were found between treatments in this study when frequencies were compared for the limited number of participants (N=27) who completed all six focal drafts. These may be the most responsible students who were able to complete successful self-edits without regard to the feedback type received, which may have masked the differences between treatments.

3. In Ferris's (2006) WCF study, the standard deviations were sometimes three times larger than the mean score. An anonymous reviewer commented that it is typical for the variability to increase with increasing language proficiency. The reviewer also noted that the higher variability in Ferris' study may be explained by the heterogeneity in the first language background of her participants as opposed to a high homogeneity of the participants in this study, all but one of whom had English as their L1.

4. Following the recommendation of an anonymous reviewer to address the high variance, all statistical analyses were repeated after removing the outliers, namely the participants whose writing showed unusually high amount of errors. This reduced the variance but did not change the results in terms of significance.

5. An anonymous reviewer pointed out that more information on how often and in what way learners misinterpret teachers' feedback as well as what percentage of errors is "untreatable" may help shed more light on this study's findings. Although it was beyond the scope of this investigation to pursue these questions, I agree that they present a very promising and necessary direction for future research.

6. I am indebted to an anonymous reviewer for highlighting the importance of teacher prose commentaries and of content-related feedback. Although the focus of this investigation is primarily on feedback on mechanical accuracy, another study is underway that examines a large variety of feedback types used in different foreign language programs.

References

Arnold, N., Ducate, L., & Kost, C. (2009). Collaborative writing in wikis: Insights from culture projects in intermediate German classes. In L. Lomicka & G. Lord (Eds.), *The next generation: Social networking and online collaboration in foreign language learning* (pp. 115-144). San Marcos, TX: CALICO.

Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12,* (3), 409-431.

Briggs, J., Di Donato, R., Clyde, M., & Vansant, J. (2008). *Workbook to accompany Deutsch, Na Klar!: An introductory German course*. New York, NY: McGraw-Hill.

Brown, H. D. (2007). *Teaching by principles. An interactive approach to language pedagogy*. White Plains, NY: Pearson Education.

Byrnes, H. (2009). Emergent L2 German writing ability in a curricular context: A longitudinal study of grammatical metaphor. *Linguistics and Education, 20*(1), 50-66.

Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, *12*, 267–296.

Di Donato, R., Clyde, M., & Vansant, J. (2008). *Deutsch, na klar!: An introductory German course*. Boston: McGraw-Hill.

Ellis, R., Sheen, Y., Murakami M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System, 36,* 353-371.

Elola, I. (2008). Portrait of a teacher: Beliefs on feedback and revision in the foreign language classroom. In H. J. Siskin, (Ed.), *From thought to action: Exploring beliefs and outcomes in the foreign language program* (pp. 48-66). Boston, MA: Thomson/Heinle.

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). Cambridge: Cambridge University Press.

Ferris, D. R. (1999). The case for grammar correction in L2 writing classes. A response to Truscott (1996). *Journal of Second Language Writing*, *8*, 1–10.

Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge: Cambridge University Press.

Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, *10*, 161–184.

Please share your stories about how Open Access to this item benefits you.

Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *Modern Language Journal, 80*(3), 287–308.

Heift, T. (2008). Modeling learner variability in CALL. *Computer-Assisted Language Learning*, *21*(4), 305-321.

Heift, T. (2010). Prompting in CALL: A longitudinal study of learner uptake. *Modern Language Journal*, *94* (2), 198-216.

Heift, T. & Rimrott, A. (2009, May). Task-induced variation by L2 learners of German. Paper presented at the 2009 conference of the Canadian Association of Applied Linguistics, Ottawa, Ontario.

High, J. L., Hoyer, J. M., & Wakefield, R. (2002). Teaching "process editing" skills with computers: From theory to practice on a larger scale. *Die Unterrichtspraxis/ Teaching German, 35*(2), 154-165.

Hyland, K., & Hyland, F. (2006). Contexts and issues in feedback on L2 writing: An introduction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 1- 19). Cambridge: 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.

Krashen, S. D. (1984). *Writing: Research, theory, and application*. Oxford: Pergamon Press.

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

Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing, 17*(2), 69-85.

Mantello, M. (1997). Error correction in the L2 classroom. *Canadian Modern Language Review, 54,* 127-131.

Melin, C. (1998). An integrative approach to process essays at the intermediate level. *Die Unterrichtspraxis/Teaching German, 31*(2), 160-168.

O'Donnell, M. (2007). Policies and practices in foreign language writing at the college level: Survey results and implications. *Foreign Language Annals, 40*(4), 650-671.

Ortega, L., & Byrnes, H. (2008). Theorizing advancedness, setting up the longitudinal research agenda. In L. Ortega & H. Byrnes, (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 281-300). New York: Routledge/Taylor & Francis Group.

Polio, C., Fleck, N., & Leder, N. (1998). "If only I had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, *7*, 43–68.

Please share your stories about how Open Access to this item benefits you.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, *20*, 83–93.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning and Technology, 13*(1), 96-120.

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

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41,* 255-283.

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

Truscott, J. (2007) The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing, 16,* 255-272.

Wible, D., Kuo, C.-H., Chien, F., Liu, A., & Tsao, N.-L. (2001). A Web-based EFL writing environment: Integrating information for learners, teachers, and researchers. *Computers and Education, 37*, 297–315.

Please share your stories about how Open Access to this item benefits you.

Appendix

Essay correction codes

Blue marks: verb mistakes

VF – verb form, e.g. infinitive (*gehen*) instead of a participle (*gegangen*)

Aux – inaccurate auxiliary verb (e.g. *haben* instead of *sein*)

Sep – separable/inseparable verb prefixes

T -   verb tense (e.g. present instead of past tense)

Ref – reflexive particle missing/unnecessary/inaccurate (e.g. *dich* instead of *sich*)

Red marks: noun and adjective mistakes

C -   case, e.g. Nominativ, Akkusativ

G -   gender, e.g. *der, die, das*

E -   endings (often adjective endings)

Purple marks: word choice mistakes

W -   problem with word choice or missing word

Prep - inaccurate/unnecessary/missing preposition

Conj - inaccurate/unnecessary/missing conjunction

Green marks: sentence organization mistakes

WO - word order (often verb position in the sentence)

Punc – punctuation (often missing/unnecessary comma)

NS – new structure needed: meaning is not clear; rewrite sentence/clause

Pink marks: spelling mistakes

Sp -   spelling, also umlauts and capitalization