An Experiment in the Relationship of Types of Written Feedback to the Development of Second-Language Writing Skills

CHRISTINE GORING KEPNER
Department of Foreign Languages
Wheaton College
Wheaton, IL 60187

ERROR-CORRECTION IN SECOND LANGUAGE (L2) students' writing is of perennial concern to L2 teachers, as is the question of whether students should be required to produce any sustained writing at all (that is, beyond sentence-level) in the target language before their oral skills are fully developed (58). Many L2 teachers fear the "fossilization" of errors (49; 59), and feel morally obligated to correct all mistakes in L2 student written work. Traditional L2 teachers may avoid requiring students to engage in sustained writing assignments, because of the burdensome task of correcting and explaining the many surface-level errors likely to occur. However, in recent years, and linked to the Writing Across the Curriculum (WAC) movement (4) interest has emerged in the value of discourse-level writing for the purpose of promoting thinking and learning (32; 39). In L2 classrooms many WAC programs have taken the form of journal-writing in the target language, even at the elementary level (13; 29; 30; 33; 37; 41; 47; 52). WAC practitioners employ a wide variety of feedback modes. Of particular concern to L2 teachers is how to respond to their students' personalized and sustained writing; should written feedback take the form of error-corrections or should it pertain more to the content of the writer's message?

PURPOSE

Research is needed which attempts to identify specific types of corrective feedback or teacher response which are most helpful for various subject matters, for different learning tasks and for differing ability levels (10). The present study sought to identify types of written feedback/response which might be related to achievement in student writing in L2 courses (specifically, college intermediate Spanish courses), and which might also be related to a low- or high-verbal ability level among student subjects. In addition to assigning two different types of written feedback to student guided-journal writing over the course of a semester (error-corrections vs. message-related comments), this study undertook to measure not only level of grammatical accuracy in student writing, but also level of thinking as indicated by the content of the journal entries. The study also sought to devise a model of written feedback which would be empirically related to the promotion of higher-level writing skills in intermediate college-level students of Spanish.

Traditionally, L2 writing achievement has been defined as mastery of the discrete surface skills required for production of an accurately-written document (9; 15; 16; 57). A logical outcome of that definition would be that L2 teachers would view it as their duty to see to it that their students evidence accuracy in grammar, vocabulary and syntax at the sentence level. Many practicing L2 teachers have continued to subscribe to this view (18; 54; 59). The present study attempted to embody this practice in the independent variable of surface error-correction. All sentence-level errors in the student writings in the surface-corrections group were identified, corrected, and explained via a brief rule or note. This explicit teaching strategy was based in theory on McLaughlin's notion of controlled processing, Omaggio's
practice of structures stage (36), and Rivers' original notion of skill-getting. (All three of these theorists/practitioners have envisioned a basic dichotomization between traditional, deductive formal-grammar L2 learning, and communicative, inductive, "natural/whole-language L2 acquisition.) The view of the learner was that of a cognitively active, aware, and curious individual who sought not only to correct but to understand the nature of his/her grammatical and lexical errors.

The alternate/experimental type of written feedback response used in this study, message-related comments, corresponded to an alternate, proficiency-related view of L2 writing. This view went beyond the sentence-level surface skills of the text to assess the more global meanings conveyed by the author. The focus was on what the writer was thinking and saying, not on how he or she was saying it. The emphasis was on communicating meaning (34; 48), and the assumption was that student learning is governed more by communicative need than by syllabus design, that the need to express meaning is the primary motivating force in language learning, and that form will arise and be acquired out of attempts at communicating (55). (This view would correspond in theory with McLaughlin's notion of automatic processing, with Ornaggio's (36) notion of communicative practice, and with Rivers' original notion of skill-using.) Thus, this feedback/response type identified no errors but, rather, responded meaningfully to the L2 student text in complete L2 sentences and paragraphs. This implicit teaching strategy was based in theory on Bialystock's notion of implicit linguistic knowledge (the intuitive, automatic information that the language learner uses in the target language), on Krashen's notion of acquisition, and on Terrell's Natural Approach. For this feedback treatment also, the view of the learner was of a cognitively engaged writer who sought to make meaning in the L2.

REVIEW OF RESEARCH

Theory and research in first-language (L1) literacy has been based during the last decade on the notion of process writing as a vehicle for the discovery and communication of meaning, which has grown out of the WAC and whole-language movements (6; 13; 14). Whole-language theory maintains that "there is no sequence of skills in language development.

Teaching kids about language will not facilitate their use of language" (13: p. 26). Process-writing teachers assume that knowledge and ability accrue as a consequence of engagement in meaningful, functional activities: "do now and learn incidentally" (51). Process-writing is, by definition, subject to formative rather than summative evaluation, and often takes the form of journal-writing as a means to reflect on new material as well as to synthesize prior knowledge and experience with new data (11; 24). Written response to process writing should be primarily formative in philosophy and in practice (21); the teacher-as-responder should function as a diplomatic coach (42) who offers reactions and advice first to content issues (11), within an atmosphere of trust (7) and a context of shared goals. Correction of discrete errors should occur only at the final stages of editing, when the piece is prepared for "publishing" or other forms of public display (11; 14).

Diaz, Kreeft, and Hildenbrand have studied the use of target-language journaling as a device for L2 acquisition. Each has found that attention to meaning rather than form characterizes the effective use of student journals within the L2 classroom. Research on the ideational or content quality of L2 student writing in relation to whole-language vs. discrete-error-correction types of feedback is sparse. However, Semke found via an experimental study that L2 student writers who received content-focused written comments on their journals spent more time writing and became more fluent than did those students whose journals received error-corrections; students who were required to correct their own errors in their journals did not display greater accuracy by the end of the semester than did those students who received error-corrections or content-focused comments. Zamel's (61) textual-analysis study of written markings made by ESL teachers on student compositions suggests that written response which combines error-corrections with positive comments regarding content or organization can only be confusing and contradictory, as students are not likely to know which type of response deserves higher priority. Zamel's examination of revised student texts revealed that the majority of revisions were on the basis of local corrections and that the teachers accepted and approved of these superficially better texts, even when they had recommended more substantive changes. Zamel recommends
that when responding to L2 student texts, meaning-level issues be considered first, "by probing, challenging, raising questions, and pinpointing ambiguities" (61: p. 96).

POPULATION AND SAMPLE

The sample consisted of sixty students drawn from four sections of Spanish 201 (Intermediate) at Wheaton College, a small midwestern liberal arts institution. The four sections were taught by two instructors, each teaching two sections. All sections utilized similar syllabi and textbooks focusing on a proficiency-based curriculum, which in addition to promoting skill-development also sought to foster the integrative themes of faith and learning. Most students were completing a general education foreign language requirement, and were either freshmen or sophomores.¹

Students were classified according to high- and low-verbal ability levels based on median splits of ranked previous English grade and ranked grade-point average (Pimsleur, 1968, found a correlation of r = .62 between L2 achievement and grade-point average, and a correlation of r = .57 between L2 achievement and previous grade in English). This data was collected unobtrusively during the first class session of the semester at which time students filled out information sheets.² Any student who failed to place in the same classification (either high or low) twice as a result of these measures did not participate in the experiment. Potential subjects were then blocked according to instructor to control for classroom effect; since there were two instructors and two verbal-ability levels, a total of thirty potential subjects from each instructor were identified, with fifteen of each instructor's group falling into the high-ability level and fifteen into the low-ability level. Four cells of fifteen subjects each were formed as subjects were then randomly assigned to one of the two feedback treatment groups by flipping a coin, resulting in a 2 x 2 factorial design, according to Kerlinger's paradigm.

VARIABLES AND INSTRUMENTATION

The active (treatment) independent variable was the type of written feedback, message-related comments vs. surface error-corrections, which appeared in written form on the guided student journal assignments. Feedback was administered to each journal entry written by the student subjects for all eight assignments, not by the students' instructors but by the researcher, who used a green pen as a writing tool. The message-related comments feedback type was whole-language in approach, responding communicatively in Spanish to the student writer and his/her message. Four criteria composed the paradigm for this independent variable of interest: 1) personalization (student's name was used); 2) summarization statement of the main point of the piece; 3) reader's reaction and evaluation of how the message was integratively communicated in light of the four elements of the assignment; and 4) a question or a suggestion for extending and/or improving upon the topic in subsequent entries (45).

The surface error-corrections treatment reflected traditional practices in L2 teaching, and as such functioned as a type of naturalistic control for the study. This feedback identified and corrected all sentence-level errors in the
journal entries in this group; additionally, reasons for the corrections were provided via a brief rule or note.

The second independent variable, verbal ability, was a fixed, classification variable. It was determined as described in “Population and Sample” above.

Two dependent variables sought to assess the effectiveness of the two types of feedback manipulated in the study. A set of student journal entries (produced during week twelve as homework) was scrutinized by the researcher and a co-rater to produce two dependent measures: a higher-level propositions count and a surface-level errors count. The higher-level propositions count was an attempt to quantify the ideational quality of the student text as well as to assess the effect of the message-related comments feedback. This measure was operationalized as tally counts of propositions or propositional clusters within the student text which exemplified the cognitive processes of analysis, comparison/contrast, inference/interpretation, and/or evaluation. This operationalization was based on Quellmalz’ (43; 44) and Stiggins et al. conceptualization of critical thinking skills, and their recommendations for designing curricula in a variety of subject matters around these four seminal cognitive processes.

The surface-level errors count was a tally total of all incidences of sentence-level mechanical errors of grammar, vocabulary and syntax within student journal texts (as in Brière). This measure sought to assess the effectiveness of the error-corrections/rule-reminders written feedback, and to measure the subjects’ level of accuracy in producing L2 text.

The two dependent measures were validated according to pilot study protocol, producing an inter-rater reliability coefficient of $r = .84$ for the higher-level propositions count and an inter-rater reliability score of $r = .92$ for the errors count. Additionally, the researcher worked with a co-rater in assessing the experimental data, resulting in an $r$ of .92 on the higher-level propositions count and an $r$ of .97 on the errors count measure.

PROCEDURES AND DATA COLLECTION

Prior to the beginning of the semester during which the study was conducted, the researcher and the two cooperating teachers developed a series of eight writing tasks for the Spanish 201 journals which would engage students in higher-level processing and induce them to integrate and synthesize learnings in at least three of the following four domains: 1) the individual student’s life experience; 2) the individual student’s spiritual journey; 3) a cultural or literary concept under study in the classroom; and 4) a grammar concept under study in the classroom. On the first day of class, students were given a journal assignment sheet which listed the dates, instructions and topics for the journal assignments (students did not become aware until the end of the semester that the journals were part of a research project). The students were told by their instructors that periodic (every two weeks) assigned journal entries, written as homework in spiral notebooks, would be required for a grade worth fifteen percent of the coursework total. Each entry was to be of no fewer than 200 words in length; grades would be based on the amount of understandable Spanish produced. Any incomprehensible language was bracketed and did not count as part of that journal entry. (The researcher spot-checked for numbers of words per entry; on average, students wrote in excess of the 200-word minimum.)

During the first week of the semester, potential subjects for the study were assigned to high- and low-verbal-ability groups according to the criteria explained under “Population and Sample.” Then subjects were randomly assigned to written feedback-type groups, to receive either error-corrections feedback or message-related comments feedback.

The journals, after being collected by the instructor and supplied with appropriate written feedback by the researcher subsequent to each assignment, were returned to students via their regular classroom teacher within one to two class sessions. At the end of the semester, after debriefing students about the study and obtaining their consent to use their data in the study, the researcher photocopied entry number six (written during week twelve) from each journal, and conducted error counts and higher-level proposition counts on each text. (Entry six was chosen as material for the dependent measures because it was surmised that by the twelfth week of the course, the independent variable of feedback type would have had ample opportunity to “work,” and the potentially confounding variable of late-semester fatigue or “spring fever” would not yet be affecting students’ performance.)
DATA ANALYSIS & RESULTS

The experimental data (the total error counts and the total higher-level statement counts) from the week twelve set of journal texts were subjected to parallel two-way factorial analysis (ANOVA) of variance in order to test the null hypothesis of no differences between groups and no interactions between levels of the verbal-ability independent variable (as in Hinkle, Wiersma, and Jurs). The means and standard deviations of the higher-level propositions counts (Table I) suggest strong main effects for both independent variables. When writing in their guided journals, students who received the message-related comments feedback produced a significantly greater number of higher-level propositions in their guided journal entries ($M = 12.1$) than did students who received the error-corrections feedback ($M = 10.4$). Two-way ANOVA score of $F = 6.509$, $df = 1, 56$, significant at the $p < .05$ level, confirmed the power of the feedback treatment variable (see Table III). Additionally, high-verbal ability subjects produced a significantly greater number of higher-level propositions in their writing ($M = 12.433$) than did the lower-verbal ability subjects ($M = 10.067$) (Table I). Two-way ANOVA (Table III) yielded an $F$-value of 12.616, significant at the $p < .01$ level, on the verbal-ability data.

Table II summarizes the surface errors-counts for both feedback-type groups and verbal-ability groups for the week twelve journal entry. No significant differences emerged for either one of the two main effects; two-way ANOVA on these data (see Table IV) produced $F$-values of 2.231 for the feedback treatment variable and of .986 for the verbal-ability variable, neither of which is significant at the $p < .05$ level. When writing in their journals, students who received the error-corrections/rule-reminders written feedback ($M = 37.87$) did not produce statistically significantly fewer errors than did students who received the message-related comments feedback ($M = 44.43$). Nor did higher-verbal ability students produce significantly fewer errors ($M = 38.967$) than did the lower-verbal ability students ($M = 43.33$).

In no case did significant interactions between the two independent variables occur (see ANOVA Tables III and IV). No statistically significant interactions emerged between feedback type and level of verbal ability. That is, the two feedback types affected both higher- and lower-verbal ability students to a similar degree. The higher-verbal ability writers consistently outperformed the lower-verbal ability writers in terms of producing a greater number of higher-level propositions in their journals; however, the higher-verbal ability writers did not produce a significantly smaller number of surface-level errors in their journals than did the lower-verbal ability writers.

**TABLE I**
Means and Standard Deviations of Higher-Level Proposition Counts as a Function of Written Feedback Treatment (Columns) and Verbal Ability (Rows) (Week 12 Journal Entries)

<table>
<thead>
<tr>
<th></th>
<th>Error-corrections</th>
<th>Message-related comments</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Low-verbal ability</td>
<td>9.6</td>
<td>2.772</td>
<td>10.533</td>
</tr>
<tr>
<td>High-verbal ability</td>
<td>11.2</td>
<td>2.242</td>
<td>13.667</td>
</tr>
<tr>
<td>OVERALL</td>
<td>10.4</td>
<td>2.608</td>
<td>12.1</td>
</tr>
</tbody>
</table>

**TABLE II**
Means and Standard Deviations of Error-Counts as a Function of Written Feedback Treatment and Verbal Ability (Week 12 Journal Entries)

<table>
<thead>
<tr>
<th></th>
<th>Error-corrections</th>
<th>Message-related comments</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Low-verbal ability</td>
<td>41.33</td>
<td>20.26</td>
<td>45.33</td>
</tr>
<tr>
<td>High-verbal ability</td>
<td>34.4</td>
<td>15.08</td>
<td>43.53</td>
</tr>
<tr>
<td>OVERALL</td>
<td>37.87</td>
<td>17.90</td>
<td>44.43</td>
</tr>
</tbody>
</table>

TABLE III
Two-Way Analysis of Variance of Higher-Level Proposition Counts as a Function of Written Feedback Type and Verbal-Ability Level (Week 12 Journal Entries)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (feedback type)</td>
<td>1</td>
<td>43.35</td>
<td>43.35</td>
<td>6.509*</td>
</tr>
<tr>
<td>B (verbal ability)</td>
<td>1</td>
<td>84.02</td>
<td>84.02</td>
<td>12.616**</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>8.82</td>
<td>8.82</td>
<td>1.324</td>
</tr>
<tr>
<td>Within cells</td>
<td>56</td>
<td>373.07</td>
<td>6.66</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>59</td>
<td>509.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; F<sub>c,v</sub> = 4.02.

**p < .01; F<sub>c,v</sub> = 7.10.

DISCUSSION & CONCLUSIONS

The data from the error-counts measures suggest that the consistent use of L2 teachers' written error-corrections combined with explicit rule reminders as a primary medium of written feedback to periodic discourse-level L2 student writing is ineffective for promoting the development of writing proficiency in the L2, whether for higher-verbal ability or for lower-verbal ability learners, at the college intermediate level. In this study, error-corrections and rule-reminders seemed to serve neither for significantly improving students' level of written accuracy in L2 surface skills nor for enhancing the ideational quality of L2 students' writing.

The data from the higher-level propositions counts measures suggest that the consistent implementation of a message-related comments model as a primary medium of written feedback to periodic discourse-level L2 writing is effective for promoting the development of writing proficiency in the L2 (in college intermediate courses), in terms of both ideational quality and surface-level accuracy. A paragraph and/or isolated sentences or phrases written by the teacher in the L2 in response to the ideas and content of the student-written piece may serve as a model of correct L2 for the student, while communicating personalized encouragement as well as higher-order questions and meaningful, content-related comments designed to promote critical thinking in the emerging L2 writer.

When taken together, combined data from the error-counts and the higher-level proposition-counts suggest that the use of a meaningful, communicative, content-related type of written feedback (as opposed to comprehensive, discrete-point error-corrections and rule-reminders) on discourse-level L2 students writing does not sacrifice accuracy for content. In this study, subjects who had received error-corrections written feedback on their L2 journal writings did not perform significantly better on post-treatment accuracy measures than did subjects who had received the message-related comments written feedback on their L2 journal entries.

Additionally, the error-corrections feedback treatment seems to have affected both verbal-ability levels to the same nonsignificant extent: There is no significant difference between the error-count scores of the group receiving consistent error-corrections feedback and the group receiving consistent message-related comments feedback. Nor is there a significant difference between the error-count scores of the low-verbal-ability group and the high-verbal-ability group.

Therefore it may be concluded that the error-corrections written feedback type is not helpful for either of the following outcomes: it does not help (to a significant degree) student L2 writers of either verbal-ability classification to avoid surface-level errors, and it certainly does not facilitate (to a significant degree) the production of higher-level writing by L2 student writers receiving that feedback consistently. The error-corrections written feedback model is helpful only in that it permits low-verbal-ability students to perform at the same level as high-verbal-ability students on measures of accuracy in L2 writing (the surface-errors count).

However, when one considers that the combined cells of low-verbal-ability subjects and high-verbal-ability subjects receiving the sur-
face errors-correction feedback wrote journal entries of significantly inferior ideational quality to journal entries of the combined ability-level cells of subjects receiving the message-related comments feedback, the one apparent virtue of the error-corrections treatment is vitiated.

In week twelve writing sample, regardless of feedback type, the higher-verbal ability writers generated significantly greater numbers of higher-level propositions than did the lower-verbal ability writers. A clear interpretation which emerges from this finding is that verbal-ability distinctions will play a significant role in achievement, particularly on learning tasks engaging higher-order cognitive processes, an inevitable challenge in the L2 classroom. It is incumbent upon teachers to recognize verbal-ability differences and to devise individualized ways of "guiding" (to use Vygotsky's term) or of "scaffolding" (to use Ausubel's as well as Bruner's term) individual learners into situations where conditions are optimal for learning to occur (Vygotsky's "zone of proximal development").

The use of writing prompts which are personally as well as cognitively engaging are essential to developing writing proficiency in emerging L2 writers. "We don't learn language by having our errors pointed out and corrected; we learn as a by-product of using language in order to do things we care about doing" (22: pp. 82-83). All subjects in this study responded favorably to a Likert-scale item assessing the students' perception of the writing topics assigned for the journals. Several wrote prose comments to the effect that "the topics were interesting and helped us to think in Spanish."

Further research on types of written feedback should attempt to determine whether written feedback is always best for all L2 student writers. Research suggests that learners respond in different and unexpected ways to teachers' written comments (17; 38; 63). Research is needed which will help identify optimally effective feedback modes which will enable L2 learners and teachers to work together toward mutual goals. Further study is needed on feedback devices such as peer response, computer-assisted writing, and individual teacher conferences, as well as on various types of written response. Learner variables other than level of verbal ability should be examined, such as locus of control. This study has demonstrated that a message-related comments type of written feedback is significantly related to the production of higher-level writing in college intermediate-level L2 student journals, and that an error-corrections type of written feedback is not significantly related to the production of more accurate Spanish in L2 students' journals. Further research on the relationship of feedback types, different L2 writing tasks, and various learner characteristics seems clearly in order.

NOTES

1 The researcher wishes to express appreciation to the Spanish section of the Foreign Language Department at Wheaton College, to the instructors of the 201 sections that participated, to the students who permitted their journal writing to be observed and analyzed, and especially to Dr. Kathleen Marshall Pederson.

2 The self-report data was later verified by checking students' academic transcripts.

3 Sample topics included the following: 1) Mis temores más nítidos; 2) Un evento memorable/crucial de mi pasado; 3) Ejemplos de la gracia en mi vida - o - Lo que yo pienso de la pena de muerte; 4) Cómo ha contribuido mi familia a la formación de mi identidad; 5) Mi opinión sobre la mujer en posiciones de autoridad y liderazgo;

6) ¿Le importa el prejuicio a Dios? — o — ¿Es el prejuicio una bendición disfrazada? — o — Maneras en que yo he experimentado el prejuicio; 7) Lo que necesita todo niño humano — o — Lo que busco en mi familia futura — o — ¿Cómo pueden los padres ayudarles a los hijos a sobrepasar la envidia?; 8) Dónde quisiera yo vivir y por qué.

Translation: 1) My most intimate fears; 2) A memorable/crucial event from my past; 3) Examples of grace in my life — or — What I think of capital punishment; 4) How my family has contributed to the formation of my identity; 5) My opinion about women in positions of authority and leadership; 6) Does prejudice matter to God? — or — Is prejudice ever a blessing in disguise? — or — Ways in which I have experienced prejudice; 7) What every human child needs — or — What I seek in my future family — or — How can parents help children overcome envy?; 8) Where I would like to live and why.

BIBLIOGRAPHY


42. Purves, Alan C. "The Teacher as Reader: An
Research Perspectives in Adult Language Learning and Acquisition, October 11–12, 1992

OPENING ADDRESS
Richard D. Lambert, Director, National Foreign Language Center
“Strategic Research: Exploring Structure and Function”

KEYNOTE SPEAKERS
Susan Gass, Michigan State University
“Second Language Acquisition: Looking Backwards, Looking Forwards”

James Noblit, University of North Carolina
“Technology and Language Learning”

Claire Krashen, University of California, Berkeley
“The Apprenticeship of Difference”

For additional information, contact:
The College of Humanities Foreign Language Center
Ohio State University, Columbus, OH 43210
Telephone: (614) 292-4361