Improving Accounting Majors' Writing Quality: The Role of Language Analysis in Attention Directing

A. Faye Borthick and Ronald L. Clark

ABSTRACT

This study investigated the role of language as an attention-directing device in improving student writing quality in an accounting systems course. Language analysis is the application of rules governing the syntax, or structure, of writing; it does not include understanding of meaning or the use of literary license. Computer programs that perform language analysis compute readability indices for the writing as a whole and create marked copy identifying potentially improvable structures, such as a preponderance of passive-voice verbs. Although they do not indicate specifically how to revise the marked structures to improve the writing, language analyzers simplify the writer's editing tasks by calling attention to sentence structures violating syntactical rules or stylistic guidelines.

In this experiment, students using a computer-implemented writing aid for language analysis scored better on two different written assignments than students not using the aid. The assignments were (1) documentation for beginning users of a microcomputer spreadsheet program and (2) a memorandum recommending a computer hardware upgrade plan for a growing company. The results were adjusted for (1) number of hours each student worked on the assignments, (2) student scores on a preliminary diagnostic test, (3) student grade-point averages, (4) student academic standing in course credit hours, and (5) student gender. The finding of improved performance with the automated aid suggests that use of writing aids for language analysis has potential for improving accountants' written communication.

ACCOUNTANTS' written and oral communication skills are of continuing concern to practitioners and educators. Several studies document (1) the importance of accountants' communication skills [AAA, 1986; Estes, 1988; S., 1990] and (2) the necessity of improving these skills [C., 1987; F., 1990; L., 1990]. The findings of this study support the view that improved language analysis is an effective tool for improving student writing quality.

A. Faye Borthick is Assistant Professor of Accounting at the University of Tennessee, and Ronald L. Clark is Assistant Professor of Accounting at The University of Georgia.
1979], (2) the gap between current and desired skill levels [Andrews and Koester, 1979; Andrews and Sigband, 1984], and (3) strategies for improving communication skills [Addams, 1981; Andrews and Pytlak, 1983; Ingram and Frazier, 1980; May and Arevalo, 1983]. Students, however, fail to recognize the importance of communication skills [Rebele, 1985], which hinders efforts to improve them.

The purpose of this paper is to determine the role of computer-implemented language analysis in improving student writing quality in accounting education. Ingram and Frazier's [1980] survey revealed substantial discrepancies between practitioner ratings of needed and displayed skills in grammar and punctuation. Use of language analysis might help students improve these aspects of their writing and thus narrow the gap between needed and displayed skills in entry-level positions.

A significant portion of new computing applications in practice and education involves word processing on microcomputers. Preparation of written materials with computer tools is already essential in many university curricula [Balkovich, Lerman, and Parmelee, 1985]. Indeed, Horn [1984] believes that microcomputer use may have its greatest educational impact in the area of writing.

Several authors [Bean, 1983; Case, 1985; Catano, 1985; Collier, 1983; Daiute, 1983; Harris, 1985; Sudol, 1985] explain the perceived benefits of word processing such as easier revision, improved appearance of the finished product, clearer visual feedback, and quicker capture of thoughts. These authors viewed word processing as a means of removing psychological and physical barriers to writing, leading to an increase in the quality and quantity of writing. These authors based their conclusions on personal experience and observation of small numbers of writers.

In addition to word processing, computer programs are available for language analysis [Frase, 1983], or the examination of writing according to rules governing the syntax (or structure) of sentences. The purpose of language analysis is to identify structural flaws, such as overuse of passive voice and incorrect punctuation. Language analysis does not address other crucial aspects of good writing such as understanding of meaning and the exercise of literary license, or taking liberties with strict grammatical forms to create just the right expression.

Frase [1983] and MacDonald [1983] proposed that computer programs for language analysis could help authors improve their writing, although they did not specify the presumed linkage between language analysis and writing improvements. The developers of one language analysis system, the Writer's Workbench, reported an enthusiastic reception for it by writers who used it. In their words, "(m)any claim their writing has improved simply because..."
the system has prodded them to think about the choices they make when writing" [Cherry and Macdonald, 1983, p. 246]. Not all authors, however, are convinced that language analysis will improve writing. For example, Oliver claimed that "the machine's quantitative-based analysis of writing style might do some students more harm than good" [1985, p. 309] because of its simplistic approach to writing analysis.

To the authors' knowledge, there has been no investigation of how this attention-directing device works or whether its use improves overall performance on written assignments. In Gingrich's [1983] field study of the Writer's Workbench, writers liked the programs' immediate feedback and found more errors using the output than without it, but there was no comparison of the overall quality of writing prepared with and without the system.

The next section explains language analysis as an attention-directing device and presents the hypothesis for investigation. The experiment is explained next, followed by a presentation of the results. The last section summarizes the findings and relates them to the potential for language analysis to improve writing quality in practice and education.

LANGUAGE ANALYSIS AS AN ATTENTION-DIRECTING DEVICE

Computer programs that analyze sentence structure are called language analyzers or readability programs. These programs are examples of a class of programs called writing aids, which are designed to improve writing quality. In addition to language analyzers, writing aids include spelling correctors, punctuation checkers, automatic thesauruses, outline generators, and index builders.

Computer programs that perform language analysis compute indices for the writing as a whole and create marked copy identifying potentially improvable structures. Although they do not indicate specifically how to revise the marked structures to improve the writing, language analyzers simplify the writer's editing tasks [Rude, 1985, p. 185] by calling attention to sentence structures violating syntactical rules or stylistic guidelines.

Feasibility of Using Language Analysis

Using a word processor makes it more feasible to use computer programs for language analysis since the text is already in machine-sensible form [Frase, 1983, p. 1885]. To the extent syntactical rules are programmable, computer programs can examine sentence structure, marking rule violations for possible revision. Good syntax improves readability by ensuring structural integrity, i.e., by helping the writer adhere to expected syntactical conventions. For example, good syntax entails following straightforward grammatical rules such as making a subject and its verb agree in number and subtle rules such as giving the appropriate antecedent for participial phrases.

Syntax is the structure of writing; semantics is the meaning. Weak syntax makes the reader work harder to understand a piece of writing. Defective structure can result in ambiguous meaning which impedes understanding.
Good writers refine syntax so that it aids rather than detracts from understanding.

The Role of Attention Directing

Broadbent [1953] asserted the importance of attention in determining response. Accepting Broadbent's view of the importance of attention, Guthrie [1959] emphasized the role of attention in conditioning the learning stimuli that control responding. He observed that attention played this role in learning theory whether the stimulus-response connection was established by associative contiguity or through reinforcement. Guthrie's rule was "what is being noticed becomes the signal for what is being done" [1959, p. 186]. This principle holds that controlling (or directing) attention is an important aspect of influencing behavior.

A behavior that accounting educators want to influence is student writing. Language analysis can be thought of as a means of directing students' attention to potentially improvable structures in their writing. For the structures it marks, a language analyzer is likely to be more effective as an attention-directing device than the course instructor because the analyzer will be more consistent and available to students on demand. According to Guthrie's theory, a way to change behavior is to bring specific stimuli to the learner's attention. In this study, the stimuli comprise language analyzer markings of sentence structures violating syntactical rules or stylistic guidelines. If the theory holds, the stimuli of language analyzer markings will be associated with improved student writing.

To examine the ability of computer-implemented language analysis to improve overall performance on written assignments, we tested the following hypothesis:

*Overall performance on written assignments will be higher for students using a language analyzer than for those not using one.*

The hypothesis does not address the issue of whether any overall performance improvement would be attributable solely to correcting sentence structure independent of changing other crucial aspects of the writing such as content and organization.

THE EXPERIMENT

The experiment was conducted in both sections (with the same instructor) of a required senior accounting information systems class during Fall quarter 1985. The systems class was chosen because (1) it already required students to complete extensive written assignments and (2) students in the class would be familiar with word processing. Two students in section 1 (not using the analyzer) and one student in section 2 (using the analyzer) dropped the course before completing any assignments, leaving 36 students in section 1 and 40 students in section 2. All students who completed an assignment were included in the analysis; the numbers of these students are shown in Table 1. The course drop rate and assignment completion rate were consistent with those in prior quarters.
Assignments

Students in both sections used a word processor\(^1\) to complete two written assignments prepared out of class: (1) documentation for beginning users of a microcomputer spreadsheet program, and (2) a memorandum recommending a computer hardware upgrade plan for a growing company. In the documentation assignment, students prepared two double-spaced pages of text explaining how to carry out simple manipulations of the spreadsheet program such as starting the program, entering data and formulas into cells, saving and retrieving the spreadsheet, and printing the contents of the spreadsheet. Students had learned to use the spreadsheet program earlier in the course.

In the memorandum assignment, students prepared two double-spaced pages of text justifying their recommended hardware upgrade plans. The purpose of the upgrade was to accommodate increased processing demands. The students were given four quarterly minimum and maximum demand forecasts across three categories of computer applications. Starting with these data, students constructed plans with the conflicting goals of minimizing cost and minimizing the number of separate installations, subject to satisfying an uncertain demand.

All students prepared identical assignments except that one section used a language analyzer\(^2\) (section 2) and one section did not (section 1). No student had prior experience with the analyzer, although some had used word processors before. Students using the analyzer were instructed to consider marked passages as potentially improvable structures, not as absolute prescriptions for changing their writing.

Use of the Language Analyzer

The language analyzer accepted text input and created marked copy. The analyzer marked complex sentences, long sentences (23 words or more), and incomplete sentences. It summarized the relative frequency of individual comments with messages noting a preponderance of sentences with multiple clauses, the absence of a reasonable number of complex sentences, and lack of variation in sentence beginnings. Within sentences, the analyzer marked words and phrases considered weak, wordy, colloquial, overused, or ambig-

---

\(^1\) Most students used the word processor in the college's microcomputer labs, although a few used other facilities. The labs were available to students 14 hours a day on weekdays and six to ten hours a day on weekends. The only limitation on student choice of a word processor was that it be compatible with the language analyzer, either supported directly or capable of producing an ASCII file.

\(^2\) The language analyzer used in the experiment was RightWriter 1.1, a product of Decisionware Corporation, Longboat Key, Florida. At the time of its selection, RightWriter was the most comprehensive language analyzer for the MS-DOS machines in the college's microcomputer labs. Other computer-implemented language analyzers are: Electric Webster, Cornucopia Software Inc., Albany, California; EPISTLE (Evaluation, Preparation and Interpretation System for Text and Language Entities) [Heldorn et al., 1982], under development by International Business Machines Corporation, Yorktown Heights, New York; Grammatik II, Reference Software Inc., Walnut Creek, California; PC-Style, Buttonware, Bellevue, Washington; Punctuation + Style, Oasis Systems, San Diego, California; STAR (Simple Test Approach for Readability), General Motors Corporation, Warren, Michigan [Rude, 1985, p. 185]; and Writer's Workbench [Macdonald, 1983], Western Electric Software, Greensboro, North Carolina. The Writer's Workbench is probably the most comprehensive writing aid system.
uous; the words and phrases the analyzer marked were those appearing in its list of such words and phrases. The analyzer also marked words considered doubly negative, redundant, or repetitious. It marked unnecessary and erroneous punctuation and all incidences of passive voice.

In addition to marking the writing, the analyzer computed four indices: a readability index, a strength index, a jargon index, and a descriptive index. The indices summarized individual markings of words, phrases, and sentences into overall evaluations.

Readability. A readability index is the level of education, stated in number of years of school, a reader needs to understand the writing. The analyzer used the Flesch-Kincaid formula to calculate the index [Flesch, 1949] where:

\[ \text{Grade level} = 0.39 \times (\text{average number of words/sentence}) + 11.8 \times (\text{average number of syllables/word}) - 15.59. \]

The index ranges from 1.0 (for the first grade) to 50.0 (unreadable). The index is based on the principle that long sentences and polysyllabic words make writing more difficult to read and understand. The readability index is based on the average sentence length and the average number of syllables per word; the more words in the sentence and the more syllables per word, the higher the index.

Shortening sentences and changing words merely to get a better readability score, however, is not likely to improve the comprehensibility of the writing. As Redish and Selzer observed, "[a] readability formula only correlates certain features with reading difficulty; the features do not cause the difficulty" [1985, p. 49]. Indeed, when sentences are causally related, combining them may improve comprehension.

Strength. The strength index measures the forcefulness of the delivery of the writing's content. Strong writing makes the point clearly and concisely. A value of 1.0 indicates strong writing; values less than 1.0 indicate weaker writing. The index, calculated by a proprietary formula, decreases with the incidence of passive voice, long sentences, wordy phrases, uncommon words, cliches, negative words and phrases, slang words, unusual abbreviations, weak phrases, and ambiguous phrases. The index is normalized by the document length.

Descriptive. The descriptive index is based on the ratio of adjectives to nouns and the ratio of adverbs to verbs. Unnecessary modifiers make writing wordy and hence hard to understand; too few modifiers make writing terse and choppy. The analyzer operated on the assumption that index values between 0.1 and 0.9 reflect normal use of adjectives and adverbs.

Jargon. The jargon index is a weighted ratio of the number of jargon words to the total number of words in a document. Jargon words are words that have been overused to the point of losing their original meaning, words known only by insiders, slang words, long phrases or words where shorter

---

* The Flesch-Kincaid formula is a U.S. Government Department of Defense standard for contractors producing user's manuals for the armed services (DOD MIL-M-38784B, April 16, 1983).
ones are more meaningful, buzz words, abbreviations, and acronyms. The analyzer identified jargon as matches between words in the text and words in its jargon list. To the analyzer, an index value of 0.0 indicated acceptable use of jargon; an index value of 0.5 or greater indicated overuse of jargon. Words formed by adding suffixes such as -ize and -ability to nouns are examples of jargon if use of simpler, more standard words would improve the writing.

Data Collection

The assignments, with no section identification, were graded by three independent graders on a zero to 100 scale. Multiple graders were used to neutralize any subjective bias of a single grader. The graders assumed the role of the course instructor, grading the assignments on an overall performance basis for content and quality. There was no separate grade for writing per se, nor did the graders examine the language analyzer output for the assignments.

As part of the assignments, students completed daily time reports showing the hours they worked on the assignments. Initial student achievement in accounting information systems was measured with a diagnostic test, given on the first day of class, consisting of 20 multiple-choice questions of varying difficulty on information systems topics. Student grade-point averages (GPAs) and academic standing were obtained from university records. Academic standing was measured as the number of university-level course hours for which a student had received credit. At the end of the course, students in the section not using the language analyzer were asked whether they knew what it was and whether they had used it.

Analysis

A general linear model (GLM) across unbalanced cells (fixed effects analysis of covariance) was used to determine the significance of using the language analyzer and the variation attributable to non-treatment sources. The dependent variable was average-score, the average of the scores of the three graders, where average-score was the average score for the documentation assignment, and average-score, for the memorandum assignment. The independent variables were:

1. Use of the language analyzer: no for section 1, yes for section 2
2. Number of hours student worked on assignment
3. Student score on preliminary diagnostic test
4. Student grade-point average (GPA)
5. Student academic standing in course credit hours
6. Student gender: male or female

The number of hours a student worked on an assignment was included to adjust for performance improvements attributable solely to a student.

---

* The graders were (1) the faculty member teaching the course, (2) a faculty member teaching accounting systems courses at another university, and (3) a third-year doctoral student in accounting systems.

* The significance tests are based on type III sums of squares, which are independent of cell frequencies and invariant to ordering of model effects [SAS, 1985, pp. 466-467].
TABLE 1
Variable Means [Standard Deviations] by Section

Panel A
Documentation Assignment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Section 1</th>
<th>Section 2</th>
<th>Both Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analyzer</td>
<td>Used Analyzer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not Used</td>
<td>$N=35$</td>
<td>$N=40$</td>
</tr>
<tr>
<td>Hours worked on assignment</td>
<td>6.8 [3.1]</td>
<td>8.6 [4.6]</td>
<td>7.8 [4.1]</td>
</tr>
<tr>
<td>Diagnostic test score (percent correct)</td>
<td>30.9 [14.2]</td>
<td>31.5 [10.5]</td>
<td>31.2 [12.3]</td>
</tr>
<tr>
<td>GPA</td>
<td>3.32 [0.36]</td>
<td>3.29 [0.40]</td>
<td>3.31 [0.38]</td>
</tr>
<tr>
<td>Academic standing (credit hours)</td>
<td>164 [32]</td>
<td>176 [31]</td>
<td>170 [32]</td>
</tr>
</tbody>
</table>

Panel B
Memorandum Assignment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Section 1</th>
<th>Section 2</th>
<th>Both Sections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analyzer</td>
<td>Used Analyzer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not Used</td>
<td>$N=36$</td>
<td>$N=37$</td>
</tr>
<tr>
<td>Hours worked on assignment</td>
<td>11.1 [10.5]</td>
<td>11.6 [4.7]</td>
<td>11.3 [6.07]</td>
</tr>
<tr>
<td>Diagnostic test score (percent correct)</td>
<td>30.7 [14.0]</td>
<td>30.8 [10.6]</td>
<td>30.8 [12.3]</td>
</tr>
<tr>
<td>GPA</td>
<td>3.30 [0.37]</td>
<td>3.28 [0.41]</td>
<td>3.29 [0.39]</td>
</tr>
</tbody>
</table>

working more hours to use the language analyzer. Results ought to be more conservative with this variable than without it.

Student score on a preliminary diagnostic test was included to adjust for entering proficiency in accounting information systems. The student grade-point average (GPA) adjusted for prior academic achievement, and the student academic standing in course credit hours adjusted for the amount of prior education. Collectively, the variables diagnostic score, GPA, and academic standing adjusted for systematic bias in academic achievement and preparation across sections. Table 1 shows means and standard deviations by section for hours worked on the assignments, diagnostic test score, GPA, and academic standing.

A separate analysis was performed for the documentation and memorandum assignments. Cronbach's [1951] alpha was used to estimate inter-grader score reliability.

Two additional analyses were used to determine how sensitive the results were to self-selection of the students in section 1 to use of the language analyzer. In one model, the students in section 1 acknowledging they used the analyzer were treated as if they had been in section 2, which used the analyzer. In the other model, these students were omitted.
TABLE 2
Performance Analysis: Using and Non-Using Sections

Panel A
Documentation Assignment

<table>
<thead>
<tr>
<th>Sources of Variation</th>
<th>R²</th>
<th>DF</th>
<th>F</th>
<th>Significance of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Model</td>
<td>0.19</td>
<td>74</td>
<td>2.72</td>
<td>0.02</td>
</tr>
<tr>
<td>Sources of Variation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyzer use: yes or no</td>
<td>1</td>
<td></td>
<td>4.40</td>
<td>0.04</td>
</tr>
<tr>
<td>Hours worked on assignment</td>
<td>1</td>
<td></td>
<td>0.18</td>
<td>0.67</td>
</tr>
<tr>
<td>Diagnostic test score</td>
<td>1</td>
<td></td>
<td>0.08</td>
<td>0.78</td>
</tr>
<tr>
<td>GPA</td>
<td>1</td>
<td></td>
<td>6.32</td>
<td>0.01</td>
</tr>
<tr>
<td>Academic standing</td>
<td>1</td>
<td></td>
<td>0.31</td>
<td>0.58</td>
</tr>
<tr>
<td>Gender: male or female</td>
<td>1</td>
<td></td>
<td>1.33</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Panel B
Memorandum Assignment

<table>
<thead>
<tr>
<th>Sources of Variation</th>
<th>R²</th>
<th>DF</th>
<th>F</th>
<th>Significance of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Model</td>
<td>0.29</td>
<td>72</td>
<td>4.47</td>
<td>0.01</td>
</tr>
<tr>
<td>Sources of Variation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyzer use: yes or no</td>
<td>1</td>
<td></td>
<td>2.83</td>
<td>0.10</td>
</tr>
<tr>
<td>Hours worked on assignment</td>
<td>1</td>
<td></td>
<td>5.29</td>
<td>0.02</td>
</tr>
<tr>
<td>Diagnostic test score</td>
<td>1</td>
<td></td>
<td>0.31</td>
<td>0.58</td>
</tr>
<tr>
<td>GPA</td>
<td>1</td>
<td></td>
<td>11.48</td>
<td>0.01</td>
</tr>
<tr>
<td>Academic standing</td>
<td>1</td>
<td></td>
<td>1.23</td>
<td>0.27</td>
</tr>
<tr>
<td>Gender: male or female</td>
<td>1</td>
<td></td>
<td>0.66</td>
<td>0.42</td>
</tr>
</tbody>
</table>

EMPIRICAL RESULTS

Explained below are the results by section and with two different adjustments for self-selecting student use of the analyzer.

By Section

The inter-grader score reliability (Cronbach’s alpha) was 0.83 for the documentation assignment and 0.89 for the memorandum assignment. As shown in Table 2, the treatment effect for the documentation assignment was significant at the 0.04 level, and for the memorandum assignment, at the 0.10 level. In both cases, performance was greater in the analyzer section. For the documentation assignment, the only other significant variable was GPA (0.01 level); for the memorandum assignment, GPA (0.01 level) and hours worked (0.02 level) were significant. (These models, with first-order interactions added, had no significant interactions.)

Table 3 shows the least-squares
means for student scores by assignment; the increase in student scores due to use of the analyzer was 2.3 points (out of 100 points) for the documentation assignment and 1.5 points (out of 100 points) for the memorandum assignment. The significance of the treatment effect and the significance of the overall models (0.02 for the documentation and 0.01 for the memorandum) support the hypothesis that use of a language analyzer is associated with improved student performance. The small $R^2$s for the overall models (0.19 for the documentation and 0.29 for the memorandum) are evidence that factors besides those included in the model are important in performance on written assignments. Language analysis is silent on such crucial matters as content, meaning, literary license, organization, and suitability for the intended audience.

With Adjustments for Self-Selection

While working in the college microcomputer labs, some students recognized that not all students had the same requirements for completing the assignments. Sixteen (44.4 percent) of the 36 students in the non-using section acknowledged knowing about the analyzer at the end of the quarter, and two (5.6 percent) of those acknowledged using it. It was not possible to deny access to the analyzer to students in the non-treatment section in the microcomputer labs. Even if lab monitors could have controlled checkout of the software diskettes, students could have shared software across sections since all students used the same labs. The purpose of this section is to reveal the possible effect of student self-selection bias for using the analyzer.

To determine the sensitivity of the results to student self-selection, the analysis was repeated with two differ-

\footnote{Least-squares means adjust cell means for unbalanced designs, giving means that would have been expected had the design been balanced.}
ent adjustments. In the first adjusted analysis, the two students in section 1 acknowledging use of the analyzer were treated as users. The overall models were still significant (0.03 level for the documentation and 0.01 level for the memorandum), but the treatment effect was only significant at the 0.09 level for the documentation and at the 0.14 level for the memorandum. The increases in least-squares means for student scores, 1.9 points for the documentation and 1.3 points for the memorandum, were less than before.

In the second adjusted analysis, the two students in section 1 acknowledging use of the analyzer were omitted. The overall models were still significant (0.02 level for the documentation and 0.01 level for the memorandum); the treatment effect was significant for the documentation (0.06 level) and the memorandum (0.13 level). The results were almost as significant as those of the original model and were somewhat more significant than those of the analysis in which the two self-selecting students were treated as users. The increases in least-squares means for students' scores were 2.1 points for the documentation and 1.4 points for the memorandum. The performance improvements were almost as large as those of the original model and were better than those of the analysis in which the two self-selecting students were treated as users.

In both adjusted models, performance improvements were not quite as great as in the original model. This result is counter-intuitive as one would expect adjusted models to yield greater performance improvements. If, however, the two self-selecting students who said they used the analyzer really made only minimal use of the analyzer, then the results of the two adjusted models are consistent with those of the original model, and the two self-selecting students belong in the non-treatment section.

Student Reaction to Using the Analyzer

Student reaction to using the analyzer was generally positive. Once they had created ASCII text, students had no difficulty invoking the analyzer to obtain on-screen or printed output, nor did they comment negatively about the added task of using the analyzer. At first, students were prone to react to each analyzer marking; students became more selective in their revisions once they realized the analyzer's limitations.

In the instructor's opinion, students were more willing to accept markings of structural weaknesses, e.g., overuse of passive voice and incorrect punctuation, from the analyzer than from the instructor. Some students acknowledged that the analyzer's output was the first indication they had ever had of grammatical weaknesses in their writing; they then worked at correcting the identified deficiencies. In the instructor's opinion, their efforts were successful.

Limitations

Self-Selection. A limitation of this study was the lack of control over self-selection of students to use the language analyzer. Insofar as the section 1 students were honest in acknowledging
their use of the analyzer, the two additional analyses show the effect of this contamination. In both adjusted models, the results were not quite as significant as in the original model.

*Time Period for Use.* Another limitation was the short time period, one academic quarter, over which students used the analyzer. Students worked on the documentation assignment in weeks four to five and on the memorandum assignment in weeks eight to nine of the ten-week quarter. During weeks six and ten, students took examinations that required written essay responses for half the point value of the examinations. Thus students received four evaluations (two examinations and the two assignments in this experiment) of their writing in one quarter. The short time period for evaluation means it is not possible to determine whether the performance improvement was lasting.

*Deficiencies of the Analyzer.* Another limitation was that the language analyzer did not always apply syntax rules correctly, especially with respect to idiomatic expressions. The competence of language analyzers is likely to improve with new versions and more powerful computers. For example, a newer version of the analyzer than the one used in this experiment supports supplemental subject-specific dictionaries and permits users to add and delete words and phrases from its proscribed lists. Analyzers would perform better if they had rules for distinguishing good use of common idiomatic expressions from faulty grammar. Another potential improvement would be for analyzers to recognize improper use of often misused words, e.g., *affect* and *effect.*

In addition to recognizing specific phrases, language analyzers might also identify subject-verb disagreements, wrong pronoun cases, noun-modifier disagreements, nonstandard verb forms, and nonparallel structures. Analyzers might let users control the sensitivity thresholds to various stylistic patterns, e.g., the number of words appearing between subjects and verbs.

The potential enhancements named here are all syntactic; semantic analysis will be much more difficult than syntactic analysis and will depend on advances in natural language processing. Given the current state of analyzer performance, however, Summers' prediction of competent reviews of student writing "by a program that looks for key words, concepts, and relationships—that *grades the paper*" [1983, p. 162]—has not been realized.

**CONCLUSIONS**

The purpose of this study was to examine the role of computer-implemented language analysis in improving student writing quality in accounting education. Based on the results of two kinds of writing assignments in an accounting information systems course, students using a computer-implemented writing aid for language analysis performed better than students not using it. This finding suggests that use of computer programs for language analysis, as an attention-directing device, has potential for improving accountants' written communication.

Although these results are not generalizable without further research, the authors believe comparable results would be obtained in other accounting
courses, i.e., for written assignments in auditing, financial accounting, managerial accounting, and taxation. The documentation assignment would be comparable to explaining a financial or managerial accounting procedure; the memorandum assignment would be comparable to a memorandum explaining the implication of an audit finding or the desirability of electing specific tax treatments for a particular kind of transaction. Improvements in students' writing should then carry over into their careers.

The magnitude of improvement in student performance is similar to that of another recent study [Borthick and Clark, 1986] involving microcomputer integration in accounting education. Studies such as these suggest that educational improvements from computer use will be incremental at best, with each additional use leading to modest gains. Over time, the aggregation of many small improvements should yield large overall effects.

REFERENCES


