

# Professors' Ratings of Language Use and Rhetorical Organizations in ESL Compositions

David Mendelsohn and  
Alister Cumming

---

To what extent do the qualities of language use and rhetorical organization in compositions written by adult ESL students affect judgements of the overall effectiveness of their compositions? Do university faculty who regularly teach ESL students attribute greater value to language use or rhetorical organization in the writing of such students? Are judgements made by ESL instructors about the value of these qualities of writing consistent with those of professors of, say, Engineering or English literature?

These questions were studied empirically

by asking 26 professors — of Engineering, English literature, and ESL — at a Canadian university to rank order, at 8-month intervals, 2 similar sets of compositions written by ESL students. These 8 compositions were modified by the researchers (cf. Freedman 1979) to possess, in a 2 by 2 factorial design, distinct combinations of effective and ineffective language use and rhetorical organization. After making the ratings, the professors were interviewed about the criteria they had used, and the difficulties they had experienced in assessing the compositions.

---

## 1. PURPOSE: EXPECTATIONS AND DECISIONS

Combination courses of three to eight months duration are offered for ESL students at many North American universities. Students in these courses are usually expected to develop their abilities to write essays of a certain quality — often defined loosely as “the level of writing required for university studies”. Such courses are distinguished from other freshman composition courses because ESL students are considered to present unique needs for learning, needs which differ substantially from those of students educated in English-speaking North America. These conditions suggest three issues necessarily underpin curriculum decision-making for ESL composition courses: (1) a sense of the quality of writing improvement desired for students to achieve; (2) the extent and kind of learning possible in the relatively short duration of one or two university semesters; and (3) the selection of an approach to instruction to meet these aims. The present study addresses one aspect of issue (1) in order to clarify how

it may impinge on issues (2) and (3). The study looks at the criteria university professors in three disciplines focus on when judging compositions written by ESL students. Is there, among three groups of faculty at one university, a common sense of the qualities of writing which is used to distinguish effective and ineffective ESL compositions? A positive answer to this question would suggest a relatively straightforward goal which could direct instruction in ESL composition courses in order to achieve the maximum, desired impact on student achievement. A negative answer, however, would imply that instruction in ESL composition might have to aim for varied student outcomes — a more likely possibility given the multiple purposes such courses are expected to fulfill, the diverse competencies ESL students bring to such learning situations, and the diversity of faculty's expectations they have to meet.

### **1.1. Perceptions of ESL Writing**

Previous research and experience suggest that appreciation of student writing varies extensively, according to such factors as raters' professional backgrounds (Follman and Anderson 1967), variation in criteria from country to country (Purves 1984), or procedures for evaluation (Quellmalz 1980, White 1984). However, little study has been devoted to understanding this phenomenon in respect to the writing of ESL university students. Studies of what ESL instructors do with student writing indicate that instructors may interpret compositions differently. Cumming (1985) found that three ESL teachers responding to a student's composition focused their attention while reading on different aspects (content, rhetorical organization or language use ) of the same text. Zamel (1985) has documented how practicing ESL instructors respond to different aspects of their students' compositions through their written comments and feedback. But research has not yet clarified whether the interpretations and judgments ESL teachers make of their students' writing correspond to those made by faculty in other disciplines. This hitherto unstudied question — at the base of the present research — is central to the relevance of ESL composition courses for the university communities they serve.

Survey questionnaires have previously been used to obtain some information. For instance, Bridgeman and Carlson (1983) asked university faculty to identify the factors they believe distinguish the writing of ESL students from that of their native-speaking counterparts. Their survey found that differences in textual characteristics (like spelling, sentence structure or vocabulary) are the features university faculty say they most commonly perceive. Kroll's (1979) is one of many studies which have surveyed faculty at universities to establish the kinds of writing tasks demanded of ESL students in their academic studies. But closer study of

how university professors interpret ESL writing is conspicuously lacking in the research literature — a fact which makes the terms, contexts and results of survey questionnaires difficult to apply to curriculum decision-making. Questionnaire research is necessarily limited to the compilation of stated beliefs; it fails to provide accurate descriptions of behaviour such as could verify what university professors actually value in the writing of ESL students.

Research which has focused more closely on university faculty's perceptions of the writing of ESL students has been restricted to a single aspect of that writing — the quality of errors. It has been established that faculty respond differently to various kinds of errors ESL students make in their writing (Vann, Meyer and Lorenz 1984) and that faculty's responses to errors can vary depending on their age and academic disciplines (Santos 1986). But as Zamel (1985) emphasizes, such research has dealt only with faculty's reactions to error types, a minor detail in the overall production of writing and the social purpose of written communication. It does not demonstrate how faculty interpret students' whole compositions or what they may value in them.

## **1.2. Perceptions of Mother-Tongue Writing**

A considerable amount of research on mother-tongue writing in English has endeavored to establish which factors most influence judgments of effectiveness in students' compositions. An early study (Diederich 1964) analyzed the holistic ratings given to three hundred compositions by fifty-three raters. Three factors were found to account for most of the variation in their ratings: (1) clarity and development of ideas; (2) errors in language use; and (3) organization of material. More recently, Breland and Jones (1984) matched holistic and analytic ratings by twenty English literature professors for over eight hundred compositions sampled randomly from a pool of 80,000 essay tests. This research found that "overall organization, use of supporting materials and noteworthy ideas were the most important factors in reader judgements", whereas "syntactic and lexical characteristics were less important influences" (p. 108). Hult (1986), in a study matching the extent of rhetorical organization to holistic ratings of 60 compositions, also found high correlations between overall, impressionistic ratings and post-hoc, tree diagrams charting thematic and rhetorical organization in the same texts.

Perhaps the most carefully designed of these studies is Freedman (1979), whose experiment provides a model for the present research. Sample essays were rewritten by researchers to represent stronger or weaker texts in four categories: content, organization, sentence structure and mechanics. Twelve raters made holistic judgments of the 96 essays,

and their judgments were later analyzed against the predetermined characteristics of the texts. Analyses of variance showed that content and organization exerted the most influence on the raters' judgments, though mechanics and sentence structure interacted significantly with organization.

Studies have also been conducted within the contexts of university programs. Freedman and Pringle (1980) — investigating the relationship of rhetorical features of writing to the grades awarded to students by their professors for papers assigned in third-year university courses — found that “development, the use of supporting detail” was “by far the most significant factor” (p. 32) in a sample of over two hundred essays from four disciplines: English literature, history, geography and biology. The researchers attribute the significance of this finding to the importance, in university curricula, of students demonstrating they have acquired transmitted knowledge. Nonetheless, other factors, such as unity, organization, style, register, vocabulary range, vividness, economy and coherence, also proved to correlate significantly with grades given to the student papers analyzed.

### **1.3. Three Important Criteria**

Conclusions reached empirically in these studies of raters' judgments center on distinctions between the importance attributed to three aspects of students' compositions: substantive content, rhetorical organization and language use. These distinctions reflect differences in the pedagogical traditions which have tended to inform what it means to write effectively for academic purposes. Concern for content is the most important element in expository writing in academic courses. Concern for the logical organization of information has long been a principle central to the teaching of rhetoric. An emphasis on the accuracy of syntax and morphology has, of course, been the mainstay of traditional grammar instruction.

ESL instruction generally excludes the teaching of substantive content, but it is clear that standards of rhetorical organization and language use figure prominently in most assessments of the writing abilities of ESL students (Jacobs et al. 1981). For this reason, the two categories of language use and rhetorical organization have been selected to serve as a basis for analysis in the present study. The category of substantive content has been controlled for rather than analyzed.

### **1.4. Research Questions**

The present research has attempted to study closely how a range of professors and instructors at one university judge the compositions writ-

ten by students of Engineering whose second language is English. In particular, this research has tried to distinguish:

- the criteria implicit in professors' judgments about the compositions of ESL students;
- potential differences in the criteria used by university faculty in three different disciplines (Engineering, English literature, and ESL);
- and the extent to which the professors' explicitly stated criteria for rating these compositions correspond to the criteria they apply, in practice, in the process of making such ratings.

The study addressed the following questions:

1. Which aspect of ESL compositions do the professors consider to be more important — rhetorical organization or language use?
2. Do professors in different disciplines prove to have different criteria for judging the compositions of ESL students?
3. Do professors in different disciplines apply the criteria they claim to use when judging the compositions of ESL students? (i.e. does what they say they do correspond to what they actually do?)

## **2. PROCEDURES: RANKINGS AND INTERVIEWS**

Eight sample compositions — all on the same topic and of comparable content (in terms of the quantity and quality of ideas expressed) — were selected from several hundred composition tests written by ESL students in the previous year as a post-admission requirement in their faculty. The eight essays were chosen to possess characteristics which conform to a 2 by 2 factorial design, juxtaposing the factors of language use and rhetorical organization at two levels (effective and ineffective). Thus the eight compositions represented:

- two compositions (A and B) with effective language use and effective rhetorical organization;
- two compositions (C and D) with effective language use and ineffective rhetorical organization;
- two compositions (E and F) with ineffective language use and effective rhetorical organization; and
- two compositions (G and H) with ineffective language use and ineffective rhetorical organization.

Initial judgments about these characteristics<sup>1</sup> of the compositions were verified by an external rater (an experienced teacher of ESL writing) rating these aspects of each text. The content in the compositions was judged to be (approximately) equivalent, and thus can be considered as a factor reliably controlled for in the design of the study. The compositions

were placed in random order and assigned random "examination numbers".

University professors and instructors whose courses regularly include ESL students were sought to rate the compositions: from the Faculty of Applied Science and Engineering (henceforth called the Engineering professors); from the Department of English Literature (henceforth called the English professors) and from the ESL Unit (henceforth called the ESL instructors). Seven Engineering professors volunteered. Nine English professors volunteered (though two did not complete the second phase of the study). Ten ESL instructors volunteered.

The eight compositions were sent to each of the raters in April 1986 with an explanatory letter. The letter indicated that the papers were from an English proficiency test. The volunteer raters were asked to rank order the essays from best to worst (i.e. from 1 to 8) and return them to the researchers. On receipt of the completed ratings, the researchers telephoned the raters to ask them to explain: (1) what they had found difficult in the task and (2) what criteria they had applied in making their judgments.

It was hypothesized that papers A and B (with characteristics of effective language use and rhetorical organization) would receive the highest ratings. Papers G and H (with characteristics of both ineffective language use and rhetorical organization) were expected to receive the lowest ratings. The purpose of this part of the study was to see what the raters from different disciplines would do with papers C, D, E, and F (with varied combinations of effective and ineffective language use and rhetorical organization). Moreover, would the criteria they said they applied to rank order the compositions correspond to the criteria they actually used?

**Figure 1. Hypothesized Configuration for Initial Rankings**

<b>LANGUAGE USE</b>	<b>effective</b>	<b>C</b>	<b>?</b>	<b>A</b>	<b>1</b>
		<b>D</b>	<b>?</b>	<b>B</b>	<b>2</b>
		<b>G</b>	<b>7</b>	<b>E</b>	<b>?</b>
	<b>ineffective</b>	<b>H</b>	<b>8</b>	<b>F</b>	<b>?</b>
		<b>ineffective</b>		<b>effective</b>	
		<b>RHETORICAL ORGANIZATION</b>			

Six months later, a second set of compositions were sent to the same raters with instructions to rank order the essays from best to worst (i.e. from 1 to 8) again. This set contained the same eight compositions with new “examination numbers” — but six of the compositions were identical to the originals, and two of the compositions had been modified by the researchers. The raters were informed that some of the compositions might be identical to those they had previously rated. Compositions A, B, C, F, G, and H were left intact (and will be referred to as compositions A2, B2, C2, F2, G2, and H2 below). Composition D was modified — changed from a text with effective language use and ineffective rhetorical organization to a text (called D2 below) with effective language use and effective rhetorical organization, like compositions A and B. Composition E was modified from a text with ineffective language use and effective rhetorical organization to a text (called E2 below) with ineffective qualities in both categories, like compositions G and H.

Figure 2. Hypothesized Configuration for Second Rankings

effective	LANGUAGE USE	C2	?	→	A2	1
					B2	2
					D2	3
ineffective	LANGUAGE USE	E2	6	←		
		G2	7		F2	?
		H2	8			
					ineffective	effective
					RHETORICAL ORGANIZATION	

These manipulations were performed (using the operational definitions described above) to verify whether tendencies evident in the behavior of the raters for the first set of rankings could be confirmed. Would composition D2 move up on the raters’ implicit scale of judgment once its organization had been improved? Would composition E2 move down on the raters’ implicit scale of judgment when its organization had been made less effective? Would any differences in the behaviors of raters from different disciplines emerge, revealing differences in their implicit criteria for judging ESL compositions? As noted above, 24 of the original 26 raters completed the second task. No interviews were conducted with the raters in this phase of the study.

### 3. RESULTS: JUDGMENTS AND CRITERIA

#### 3.1. Rank Ordering of the Compositions

Mean rank orderings of the compositions were calculated (using the SPSSx Reliability procedure and the Kendall Test of Concordance) for the total group of raters and for each subgroup defined by their professional discipline. In each instance, chi square tests (at 7 degrees of freedom) showed very high levels of significance for the rankings ( $p < .00001$ ). Computation of Kendall's coefficient of concordance ( $\omega$ ) demonstrated that the levels of agreement on the rank orderings were high among the overall group of raters (.8 for the initial and second rankings) and even higher within the subgroups in each discipline: English professors'  $\omega = .9$  and  $.8$ ; Engineering professors'  $\omega = .9$  and  $.9$ ; ESL instructors'  $\omega = .8$  and  $.9$  — for the first and second ratings respectively.

Figure 3 presents the mean rank orderings for each of the eight compositions by the total group and the three subgroups of raters. As this figure indicates, the ratings conformed closely to the hypothesized configuration presented in Figure 1. For all groups (except one case with the Engineering professors), papers A and B received the highest mean ratings, papers G and H received the lowest mean ratings, and papers C, D, E, and F received various ratings in the range between. The exception made by the Engineering professors was to rank paper F (which had ineffective language use and effective rhetorical organization) at 6.9 — on a par with papers G and H (which had ineffective characteristics in both categories). In contrast, the ESL instructors rated paper F at 3.3 and the English professors rated it at 4.1. Overall, there was no clear priority given to papers C and D or E and F, except for Engineering professors who consistently ranked papers with effective language use (C and D) above papers with effective rhetorical organization (E and F). Otherwise, each paper in the middle range received inter-related rankings which did not indicate that greater or lesser values were being attributed to the characteristics of their texts by the groups.

Figure 4 shows the mean rank orderings for the second set of compositions. Again the ratings conformed closely to the hypothesized configuration (Figure 2) with little variation between or among the subgroups of raters. As predicted, adjusting the features of rhetorical organization in papers D2 and E2 led to changes in their rank ordering by all groups of raters. D2, with improved organization, moved up about one point on the rank orderings. E2, with its organization made worse, moved down about one point on the rank orderings. However, paper F2 was again treated by the Engineering professors in a way which differs substantially from the English professors and ESL instructors. The Engineering professors



Figure 3. Initial Rankings of Compositions

		Paper	Total	English	Engin.	ESL	Paper	Total	English	Engin.	ESL
<b>effective</b>	<b>C</b>		4.0	3.9	3.1	4.6	<b>A</b>	1.1	1.0	1.1	1.2
	<b>D</b>		5.2	5.2	4.3	5.7	<b>B</b>	1.9	2.1	1.9	1.8
<b>ineffective</b>	<b>G</b>		7.1	7.1	7.6	6.8	<b>E</b>	5.1	5.3	4.9	5.1
	<b>H</b>		7.1	7.2	6.3	7.5	<b>F</b>	4.5	4.1	6.9	3.3
		<b>ineffective</b>					<b>effective</b>				
		<b>RHETORICAL ORGANIZATION</b>									

Total raters: n=26, w=.8  
 English professors: n=9, w=.8  
 Engineering professors: n=7, w=.9  
 ESL instructors: n=10, w=.9  
 p>.00001 for all group rankings, chi square at 7 d.f.

Figure 4. Second Rankings of Compositions

	LANGUAGE USE					RHETORICAL ORGANIZATION				
	Paper	Total	English	Engin.	ESL	Paper	Total	English	Engin.	ESL
effective	C2	4.2	4.1	3.9	4.5	A2	1.2	1.1	1.1	1.3
						B2	1.9	1.9	2.0	1.8
						D2	3.7	3.9	3.0	4.1
ineffective	E2	6.4	6.6	6.1	6.4					
	G2	7.0	7.3	7.1	6.7	F2	4.7	4.3	6.2	3.9
	H2	7.0	6.9	7.3						

Total raters: n=24, w=.8

English professors: n=7, w=.9

Engineering professors: n=7, w=.9

ESL instructors: n=10, w=.8

p>.00001 for all group rankings, chi square at 7 d.f.

ranked paper F2 at 6.2, thus rating it among papers E2, G2, and H2 with ineffective language use and rhetorical organization. In contrast, the English professors ranked the same composition at 4.3 (almost on par with papers C2 and D2), and the ESL instructors ranked it at 3.9 (above papers C2 and D2).

### 3.2. Retrospective Interviews

Results of the telephone interviews with the raters were analyzed by *post hoc* classifications of the raters' statements. Of particular interest were statements they made about the criteria they believed they had used to judge the compositions. These statements were first categorized on the basis of whether the raters had claimed to attribute more or less importance to language use or rhetorical organization. As Figure 5 shows, all of the Engineering professors said they considered language use to be of primary importance. The English professors were split 50%-50% on this, half indicating they gave priority to each category. Among the ESL instructors, five claimed to attribute more importance to rhetorical organization, whereas two stated language use had greater importance in their

Figure 5. Criteria Stated by Raters in Interviews

		English n=9	Engineers n=7	ESL n=10
Criterion said most important	language:	50%	100%	29%
	organization:	50%	0	71%
Criterion judged most important	language:	38%	83%	0
	organization:	38%	0	71%
	both:	25%	17%	29%
Consistency between statement and judgment		63%	83%	57%
N not analyzable		1	1	3

judgments of the compositions. In all, five of the recorded statements could not be analyzed because they did not conform readily to dichotomous divisions between the two categories of language use and rhetorical organization. Most of the raters said they had little difficulty ranking the highest and lowest papers, but they found ranking the papers in between was troublesome.

The categorized statements from individual raters were next matched against the rankings of the compositions each person had made for the first set of compositions. Row 3 of Figure 5 shows the extent of consistency between criteria stated by the raters in each subgroup and the criteria they had apparently applied in their rank orderings, according to the predetermined characteristics of the eight compositions. The Engineering professors showed the greatest consistency (83.3%) compared to the English professors (62.5%) and the ESL instructors (57.1%). However, matches between interview statements and rating behavior could not be made for a further five subjects because their stated criteria did not adhere strictly to a dichotomous division between the two categories of language use and rhetorical organization.

#### **4. DISCUSSION: CONSISTENCIES AND DIFFERENCES**

Though results of this study are not conclusive, they indicate how answers can be tentatively formulated for the three questions which guided the research. Answers to the first two questions can be conflated into one, with certain qualifications: 1. Which aspect of ESL compositions do the professors consider to be more important — rhetorical organization or language use? 2. Do professors in different disciplines prove to have different criteria for judging the compositions of ESL students? There was a high degree of concordance among the overall ratings given to the sample compositions, but this agreement among raters appears to center on the extreme ends of the writing abilities represented in the sample compositions. For the middle range of ESL writing ability represented, it appears that the Engineering professors who participated in this study applied different criteria than the English professors and ESL instructors: the Engineering professors attributed more importance to language use than rhetorical organization to judge the effectiveness of compositions in the middle range of the sample compositions. Conversely, ESL instructors seem to have attributed more importance to rhetorical organization in rating the same compositions. English professors did not seem to be biased in either direction.

An examination of the overall ranking by the university faculty in the three different disciplines might lead one to the conclusion that the criteria for judging these compositions were very similar. Indeed, the results of

manipulating features of rhetorical organizations in papers D2 and E2 demonstrate that rhetorical organization is a salient criterion for all groups of the raters sampled. However, a closer examination of the middle range of the writing represented suggests different interpretations are evident among groups of raters. Somewhat different criteria were applied by different faculty groups when they rated compositions which had varying shades of such qualities as language use and rhetorical organization. That is, for ESL compositions where language use and rhetorical organization are both distinctly effective or ineffective, it appears that raters from different disciplines can readily agree on the value of such writing. But where the effectiveness of language use or rhetorical organization varies in students' texts, raters from different disciplines may attach greater or lesser importance to either one of these features. This phenomenon is most evident in the case of composition F and F2, a writing sample with effective rhetorical organization and ineffective language use. This essay was consistently rated highly by ESL instructors, lowly by Engineering professors, and given middle ranking by English professors. This is, however, a single case. Because the present study used a small number of compositions — so as to not overburden the abilities of the raters to make rank orderings (Miller 1967, Quellmalz 1980) — it is not clear whether this finding would be verified over a larger sample of ESL compositions.

Nonetheless, it is clear that a simple dichotomy between language use and rhetorical organization does not suffice to capture the complexity of judgments the raters used to rank order the ESL compositions. Several results indicate that these two criteria interacted with other criteria to form the judgments of the raters. In the first set of ratings none of the groups of raters ranked both compositions C and D (with effective language use and ineffective rhetorical organization) over both compositions E and F (with ineffective language use and effective rhetorical organization), or vice versa. Rather, the actual rank orderings of these compositions were at comparable, interrelated levels. This suggests that the predetermined characteristics of the texts were either (1) balanced out against one another by most raters or (2) they were just not the only, salient criteria the raters attended to. Likewise, in the second ratings, neither the English professors nor the ESL instructors made distinctions between these four compositions such as would indicate that the categories of language use or rhetorical organization were paramount in forming their decisions. It would thus be prudent to consider the rating of ESL compositions to be like other complex forms of decision-making, where a multiplicity of factors may enter into the process of making single judgments (Pitz and Sachs 1984).

Answers to the third research question are less clear, mainly because of

limitations imposed by the classification of data into two dichotomous categories. This question asked, do professors in different disciplines apply the criteria they claim to use when judging the compositions of ESL students? Comparison of the *post hoc* analyses of interview data with the actual rankings of the compositions shows that the Engineering professors were obviously consistent in following through with their stated conviction that language use was of greater value in informing their judgments than rhetorical organization was. This may bear out a finding from Bridgeman and Carlson's (1983) survey, where Civil Engineering professors were the only group of faculty at several American universities who claimed not to perceive the organization of ESL students' writings as a characteristic distinguishing it from that of native-speaking students of English.

The other groups of raters in the present study expressed less distinct beliefs about the importance of any one criterion in shaping their judgments. Therefore, it is not clear how their stated beliefs correspond to their behavior when judging the compositions. The ESL instructors tended to state that rhetorical organization was a more important criterion informing their judgments, but their rankings of the compositions show this was only marginally evident in practice (i.e. for one composition in the second ratings). This might be explained in terms of what they believe they "should" say differing from what they do in practice. The English professors did not claim to value one aspect of writing over another, nor did their ratings of the compositions indicate they used only one or two criteria in their judgments. Faculty who regularly teach composition may not have common, fixed criteria they apply to evaluate writing by ESL students, preferring instead to respond to the particular characteristics of individual student texts. As previous studies reviewed above suggest, a threefold interaction between content, language and organization may better capture the criteria such instructors value in student writing.

## **5. IMPLICATIONS: AUDIENCES AND AIMS**

### **5.1. Curriculum Implications**

It may be propitious that greater differences did not emerge among the judgments of ESL compositions by different groups of faculty, though more substantial differences might be found through study of broader contexts (beyond the reading of student compositions tests) and scope (beyond three groups of faculty at one university). Great differences would have signalled a discrepancy in expectations for writing by university faculty of the kind recently documented for high school teachers in different subject areas (Applebee 1984). The present small study does,

nonetheless, establish that certain differences do exist between the criteria professors in some disciplines apply to judge ESL writing. These differences have consequences for curriculum policy in ESL composition courses — to the extent that they bear on determination of the goals and focus of such instruction. Should ESL composition instruction aim to fulfill the expectations for student writing of ESL instructors, English professors, or Engineering professors? Which authority should ESL learners aim to satisfy in their efforts to make improvements in their writing? If differences exist among these faculty groups, is it incumbent upon or indeed possible for language faculty to alter the perceptions of their peers in other academic disciplines, or for language faculty to modify their own beliefs to accommodate the views of the university community at large?

These questions are vital because, as Bereiter (1980) demonstrates, students' writing development in educational settings is fundamentally determined by the writing that is required of them in their courses. ESL students, like most other students, write mainly to the demands instruction places on them, having few opportunities or needs to write for purposes or values other than those created in the contexts of their courses. If anything like a developmental progression occurs in learning to write in a second language, it is probably something which takes place over a period of many years in response to varied social contexts and students' accumulating knowledge (Mohan and Lo 1985).

The instruction which takes place in ESL composition courses necessarily has to focus on limited and clearly defined goals. The time allotted for learning in ESL composition courses is usually short, meaning that even the best of students may only be able to accomplish a few distinct achievements in the period of one writing course (Cumming 1986, Anderson 1982). The choice of relevant, attainable goals for writing improvement in ESL composition courses is thus consequential for students and instructors alike. An empirical question which still needs to be answered concerns the amount of improvement that can be achieved in ESL students' organization and in language use in the period of a short composition course.

## **5.2. Implications for Evaluation**

This study also has implications for the evaluation of the writing of ESL students. The differences which were exposed among faculty's ranking of the sample compositions suggest that purely holistic ratings of ESL compositions would have concealed variation in raters' implicit criteria for judgment, variation which would have been consequential for rating the middle range of ESL writing ability represented. Studies such as Quellmalz (1980) show how raters' judgments can vary in a number of ways in the

process of evaluating writing, but raters' attributing different values to categories like language use or rhetorical organization might be an especially salient factor in rating the writing of ESL students. Therefore, it would seem preferable in evaluation practice to use analytic scales requiring raters to identify their assessments under different categories, as in Jacobs et al.'s (1981) ESL Composition Profile. Likewise, it seems clear that the matter of who evaluates the writing of ESL students is important in evaluation practice and research. Perhaps the common significance attributed to certain characteristics of student writing in studies like of those of Freedman (1979) and Breland and Jones (1984) can be explained by their use of raters who were all English professors — who, as in the present study, proved to have quite similar criteria for assessing writing.

The quasi-experimental approach taken to this study was, with the support of post-session interviews, able to establish what some of the elements are that inform different faculty's interpretations of the compositions of ESL writing. However, different approaches to research of this issue would be worthwhile in the future. For instance, the decision to combine ordinal data (rank orderings) with categorical data (of the characteristics of the compositions) ruled out the possibility of using multivariate statistics to analyze the data. More illuminating understanding of the psychological processes of interpretation which different faculty attend to while reading ESL compositions could be obtained by eliciting concurrent verbal reports from volunteer raters with different backgrounds. As in Cumming's (1985) study of ESL teachers responding to students' texts, think aloud protocols could be expected to establish more precisely what different faculty pay attention to in students' texts. Finally, research (like Schmidt 1981) is needed on the contexts in which ESL university students write — in order to know more about the conditions under which their writing is usually performed and judged by their professors.

#### FOOTNOTES

1. Operational definitions of high or low levels of rhetorical organization were the presence or absence of an explicit method of logical ordering such as enumeration or comparison, discourse markers, paragraphing, and excessive redundancy in content. Operational definitions of language use were inaccurate and inappropriate lexis, syntax, morphology, spelling and punctuation at the average frequency of three such items per sentence (for low levels) and one or less per sentence (for high levels).

---

#### REFERENCES

- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(2), 369-406.
- Applebee, A. (1984). *Contexts for Learning to Write*. Norwood, N.J.: Ablex.



- Bereiter, C. (1980). Development in writing. In *Cognitive Processes in Writing*. L. W. Gregg and E. R. Steinberg (Eds.). Hillsdale, N.J.: Erlbaum.
- Breland, H. M. and R. J. Jones. (1984). Perceptions of writing skills. *Written Communication* 1(1), 101-119.
- Bridgeman, B. and S. Carlson. (1983). *Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students*. TOEFL Research Report No. 15. Princeton, N.J.: Educational Testing Service.
- Cumming, A. H. (1985). Responding to the writing of ESL students. In *Patterns of Development*. A. Pare and M. Maguire (Eds.). Ottawa: Canadian Council of Teachers of English.
- Cumming, A. H. (1986). Intentional learning as a principle for ESL writing instruction: A case study. *TESL Canada Journal, Special Issue 1*, 69-83.
- Diederich, P. B. (1964). Problems and possibilities of research in the teaching of written composition. In *Research Design and the Teaching of English*. Champaign, Ill.: National Council of Teachers of English.
- Follman, J. C. and J. A. Anderson. (1967). An investigation of the reliability of five procedures for grading English themes. *Research in the Teaching of English*, 190-200.
- Freedman, A. and I. Pringle. (1980). Writing in the college years: some indices of growth. *College Composition and Communication* 31(3), 311-324.
- Freedman, S. W. (1979). How characteristics of student essays influence teachers' evaluations. *Journal of Educational Psychology* 71(3), 328-338.
- Hult, C. A. (1986). Global marking of rhetorical frame in text and reader evaluation. In *Functional Approaches to Writing: Research Perspectives*. B. Couture (Ed.). Norwood, J.J.: Ablex.
- Jacobs, H. L., S. A. Zinkgraf, D. R. Wormuth, V. F. Hartfiel and J. B. Hughey. (1981). *Testing ESL Composition: A Practical Approach*. Rowley, Mass.: Newbury House.
- Kroll, B. (1979). A survey of the writing needs of foreign and American college freshmen. *ELT Journal* 33(3), 219-227.
- Miller, G. A. (1967). The magical number seven, plus or minus two: some limits on our capacity for processing information. In *The Psychology of Communication*. New York: Basic Books.
- Mohan, B. and W. Lo. (1985). Academic writing and Chinese students: transfer and developmental factors. *TESOL Quarterly* 19(3), 515-534.
- Pitz, G. F. and N. J. Sachs. (1984). Judgment and decision: theory and application. *Annual Review of Psychology*, 35, 139-163.
- Purves, A. (1984). In search of an internationally-valid scheme for scoring compositions. *College Composition and Communication* 35(4), 426-438.
- Quellmalz, E. (1980). Problems in stabilizing the judgment process. Center for the Study of Evaluation Report No. 136. Los Angeles: University of California.
- Santos, T. (1986). Professors' reactions to the academic writing of non-native-speaking students. Unpublished Paper Presented at the 20th Annual TESOL Convention, Anaheim, Calif., March 4.
- Schmidt, M. F. (1981). Needs assessment in English for Specific Purposes: The case study. In *English for Academic and Technical Purposes*. L. Selinker, E. Tarone and V. Hanzelli (Eds.). Rowley, Mass.: Newbury House.

- Vann, R., D. E. Meyer, and F. O. Lorenz. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly* 18(3), 427-440.
- White, E. (1984). Holisticism. *College Composition and Communication* 35(4), 400-409.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly* 19(1), 79-101.

## **THE AUTHORS**

### **David Mendelsohn**

David Mendelsohn is a lecturer in ESL, and Academic Co-ordinator of the TESL Certificate programme at Woodsworth College, University of Toronto.

### **Alister Cumming**

Alister Cumming is a Senior Research Officer and Ph.D. candidate at the Ontario Institute for Studies in Education.