Evaluating ESL: Making sense of university professors' responses to second language writing

Felicia Roberts [a,1], Tony Cimasko [b,*]

[a] *Department of Communication, 2170 Beering Hall, Purdue University, West Lafayette, IN 47907, USA*
[b] *Department of English, 500 Oval Drive, Purdue University, West Lafayette, IN 47907, USA*

Abstract

This study addresses the response of social science and engineering science faculty to a naturally occurring sample of second language writing. Using a matched-guise protocol, faculty participants were led to believe that the one-page essay was produced by an international student whose first language was either Chinese or Spanish. The faculty evaluated the writing holistically (on a scale from 1 to 10) and were invited to "correct the five most troublesome errors." Results indicate that the ethnic guise did not affect holistic scores; however, the social science and engineering science faculty did rate the composition differently. While qualitative analysis demonstrates that, not surprisingly, individual editing styles among faculty are quite variable, there was a tendency across faculty to edit semantic gaps as opposed to grammatical items. This indicates a preference by the faculty to clarify content, a finding that supports prior research.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Error gravity; Reader response; Writer identity; English for Academic Purposes

Introduction

Second language (L2) composition has gained significant autonomy, both as a research discipline and as an educational program (Zamel, 1995). The value of this focused attention is an ever-increasing understanding of the processes and products of L2 writing. At the same time, across university campuses, L2 writers strive to meet the everyday challenge of preparing texts for courses throughout the curriculum, and, by extension, facing the writing demands and preferences of the instructors of those courses. In studies of non-nativeness in L2 student writing, native/non-native identity has tended to be viewed only from the perspective of writers and texts

with little or no consideration of audience, or in cases when audience response has been considered, limiting research to relationships between non-native instructors and native learners (cf., Aigner & Thum, 1986; Clayton, 2000; Norris, 1991). The current study, however, examines responses to non-nativeness on the part of one of the most important audiences L2 writers encounter during their educational experience: their professors. The study builds on prior work concerning reader response to L2 writing by examining several facets of the evaluation process in a university setting: the effect of writer identity, especially as related to potential ethnic biases; evaluator characteristics; and, finally, orientation by university professors to editing errors in L2 writing.

Many approaches within second language research have presented insights into the intricacies of native/non-native interactions; one of those has been broadly termed the study of ''error gravity'' (for reviews, see Eisenstein, 1983; Ludwig, 1982; Rifkin & Roberts, 1995). This type of research has aimed to establish hierarchies of error types so that language teachers might focus on areas of grammar and pronunciation judged by native speakers to be most disruptive to communication. Measures of comprehensibility, acceptability, and naturalness are among the constructs used to assess sensitivity to particular L2 errors.

Using an error evaluation approach, error gravity hierarchies have been investigated for second and foreign language classes in English (Johansson, 1978; Khalil, 1985; Sheorey, 1986; Santos, 1987; Tomiyana, 1980); French (Magnan, 1983; Piazza, 1980); German (Delisle, 1982; Politzer, 1978); and Spanish (Chastain, 1980; Guntermann, 1978; Gynan, 1985). In most of this research, isolated spoken or written sentences or contrived prose passages have been used as stimuli. Such studies have provided useful insights into native speaker perception of errors in L2 writing, but the approach, which often presents only one error per sentence carrier, may only partially capture a realistic evaluation of L2 performance. In order to help composition instructors better prepare their learners, we argue for the importance of assessing response to L2 error in a more naturalistic manner. Rather than asking native writers to evaluate L2 errors in terms of abstract concepts such as acceptability or comprehensibility, a naturalistic performance by the evaluator (i.e., marking and editing any error they perceive) holds the potential to produce a more realistic accounting of response to L2 writing. In the current study, we aim to more closely approximate actual evaluation of and response to L2 writing by asking university faculty to holistically grade and then edit an actual composition written by a student for whom English is not a native language.

Along with investigating the way errors are responded to by university professors, we also examine whether the identity of the language *producer* influences evaluation of the language sample. This purpose brings the current study into the domain of social psychological research regarding language and ''interpersonal perception processes,'' or the ways in which individuals perceive and evaluate other individuals or groups (Bradac, 1990; Brown, Giles, & Thakerer, 1985). Much of this work, which spans several decades, has been carried out in studies of accentedness or dialect, which rely on spoken language data to elicit evaluations of non-standard speakers (for a review, see Bradac, Cargile, & Hallett, 2001). Nevertheless, it is not necessary to have speech cues to elicit attitudinal responses to individuals and groups. When asked to simply ''think about'' an ethnic or racial group (Callan & Gallois, 1983) or even look at pictures of ethnic group members (Fishman, Rattner, & Weiman, 1987) and then rate members of that group according to a list of personality attributes, subjects behave in a predictably negative or positive manner. Clearly, stereotypes exist, and they feed into attitudes and evaluations of groups and individuals.

Investigations of production in L2 writing, however, have only minimally probed whether or not pre-existing stereotypes contribute to the evaluation of errors made by non-native language

users. Notably, Rubin and Williams-James (1997) examined attitudes towards L2 writing errors by modifying essays composed by native writers for a statewide proficiency test for college juniors. The researchers included six kinds of errors and tagged the essays with pseudonyms identifying the ''authors'' as Anglo-American, Danish (intended to more broadly represent attitudes towards Northern Europeans), or Thai (broadly representing East Asians). Native mainstream English composition instructors rated the essays, noting errors and commenting briefly. Although there was no variation in the number of errors found by the teaching assistants, the researchers found that scores on the essays favored the ''Thai'' writers over both their ''Danish'' and ''American'' counterparts. Confirming earlier findings by Janopoulos (1992), comments on the essays were more frequent on the American essays than on the others, an inverse relationship suggesting more feedback for American students. The authors concluded that the raters approached native writers with a ''they-should-have-known-better'' attitude, while at the same time perhaps compensating for negative attitudes towards those more clearly separated from an English language tradition—the Asian writers (pp. 148–150).

In the current study, we focus on two prominent L2 groups on college campuses, those for whom Chinese or Spanish is a first language (L1), to ascertain if there is any evidence of differential response to L2 writing based on these cultural identities. Because we are primarily concerned with ethnic bias and seek to control for linguistic distinctions, we use a matched-guise procedure (Lambert, Hodgson, Gardner, & Fillenbaum, 1960) which allows for focused exploration of error gravity and writer identity.

Beyond L2 writing, additional research has suggested the importance of understanding the interactions of identity and error gravity beyond the classroom and the need to understand it in academic contexts in preparation for post-graduation realities. Occasional research conducted over the past quarter-century (e.g., Beason, 2001; Hairston, 1981; Leonard & Gilsdorf, 1990; Ranelli & Nelson, 1998; Williams, 1981) has shown that L1 writers in professional settings who commit relatively minor writing errors that do not impact audience understanding are nevertheless perceived as careless and not adequately dedicated. A representative example comes from Beason (2001), who offered copies of college junior writing projects with syntactical and spelling errors intact to finance and business managers in two regions of the United States for their commentary. Although the intensity of the managers' reactions in this case was variable, their consensus that the mistakes constructed ''negative ethos'' for the writers (pp. 46–58) highlights the crucial role that even minor aspects of text presentation contribute to readers' evaluations.

While the framework for this investigation draws from studies of both native writers and of second-language learners, we recognize with Silva (1993) that first language and second language writers differ on several dimensions, even if many similarities are to be found on broader levels.[2] It is for this reason that our investigation focuses exclusively on ESL writing as a distinct phenomenon on most university campuses. Indeed, the somewhat more recent identification of some students as ''Generation 1.5'' (those who are born into immigrant families but may be schooled extensively in both their home country and North America; see Harklau, Losey, & Siegal, 1999) indicates that the complexity of response to non-mainstream students will need to continue growing in response to increasing diversity.

---

[2] Silva's (1993) review and synthesis of literature comparing L1 and L2 writers pointed to broad similarities in their respective composing processes, but also found a number of important differences. These differences included slower writing and less lexical richness on the part of L2 writers, but also greater ability to identify and rectify certain categories of errors than their L1 counterparts.

Table 1
English language writing response literature

Reference	Data and procedure(s)	Participants	Conclusions
English as a second/foreign language			
Johansson (1978)	Reviews of original student essays with errors in place for subjective quality and the presence of "overt" and "covert" errors	Swedish university-level English learners English NS university students	Subjective evaluations often correlate with objective assessments; only "overt" errors have a significant impact on judgments
Tomiyana (1980)	One personal essay and one exam essay intentionally rewritten with key omissions and errors, which the participants attempted to correct Answers to questions on original writers' likely academic achievement	English NS graduate students	Article errors and omissions were easier to correct and thus did not substantially affect understanding
Vann et al. (1984)	Questionnaires: information on faculty backgrounds and isolated sentences with common ESL learner errors	Faculty from 60 academic departments	Age and academic discipline of faculty may be important to response
Khalil (1985)	Compositions written by Arab freshmen, chosen for grammatical and semantic deviations	English NS university students from multiple academic areas	Semantically deviant utterances were less intelligible; context did not influence ability to interpret; no association between intelligibility and interpretation
	Likert-scale rankings of various compositions for intelligibility and interpretability		
Sheorey (1986)	Responses to 20 decontextualized sentences representing common errors	US native ESL teachers Indian non-native ESL teachers	Native teachers tolerated errors more than their non-native counterparts; both groups viewed verb-related errors as most significant

Santos (1987)	Two student compositions displaying errors in five grammatical categories	Evaluating professors University students	Significant differences in response to errors between more complex marked forms and more basic unmarked forms
Santos (1988)	400-Word compositions, scored on content and language	Humanities and physical sciences professors East Asian university students	Humanities professors and more experienced professors demonstrated greater leniency; content was graded more harshly
Janopoulos (1989)	NNS in-class writing samples, divided into 4 proficiency categories	12 ESL native speaker composition instructors	Recall of writing sample details was greater for higher proficiency writing, demonstrating that NS holistic raters attend to meaning
Janopoulos (1992)	24 sentences displayed separately, each containing one of 12 common errors Placement examination essays	University faculty from various departments across the curriculum Intermediate-high NNS who had finished an intensive English program	Faculty were more tolerant of NNS errors than those made by NS writers Lexical variation and variation minus error correlate with high scores
Rubin and Williams-James (1997)	Proficiency test essays, re-labeled with English, Danish, and Thai surnames	Graduate NS teaching assistants NS college juniors	Poor writing correlates with more extensive comments; fewer comments for "foreign identity" NNS
Response in non-academic settings			
Hairston (1981)	Questionnaire: level of irritation to 65 decontextualized sentences with one standard usage error	Respondents from 63 non-academic occupations	Males were less bothered than females; the most serious errors suggested lack of educational status
Leonard and Gilsdorf (1990)	Decontextualized samples of common errors	Post-secondary business communication teachers Executive vice presidents	Sentence structure errors are most distracting, while lexical errors are least distracting

Table 1 (*Continued*)

Reference	Data and procedure(s)	Participants	Conclusions
Ranelli and Nelson (1998)	Survey of writing attitudes among pharmacy students	Second-year pharmacy students	Communicative and mechanical competence correlated significantly with writing ability
	Holistically scored position papers written by pharmacy students, 2/student	Unidentified readers, following an Educational Testing Service scoring guide	
Beason (2001)	A business document rewritten to reflect various error types Quantitative questionnaires indicating extent of irritation with errors Interviews with participants	Private sector managers and executives	Quantitative rankings revealed diverse range of priorities; interviews uniformly point to construction of ''negative ethos,'' including lack of care and meticulousness

The present study is intended to build on and synthesize various aspects of research on writing response (described above and summarized in Table 1). The aim is to build a more complete picture of the interaction of student and instructor identity and error perception in relation to texts and evaluation tasks intended to reflect a greater degree of authenticity. Specifically, we address the following questions:

1. Are correlations observed between evaluator responses and the assumed ethnic/language backgrounds of non-native writers?
2. Are correlations observed between evaluator responses and evaluator characteristics?
3. What qualities are observed in instructor responses to errors in non-native compositions?

Procedure

Presenting the composition in two guises

To investigate factors contributing to evaluations of ESL writing, a project was undertaken to gather faculty response to a composition written by a student enrolled in an intermediate level of intensive ESL. Because we were concerned with gathering responses from faculty across disciplines, we chose an actual L2 composition on the topic of transportation in a foreign city. This was a topic that could be relevant to both science/engineering and social science professors. The finished writing sample was about one page in length; although this does not reflect the longer, high-stakes writing expected of students across the university curriculum, it does approximate shorter reflective writings that students are asked to do more frequently, particularly in their first year (Leki, 1995, p. 253). This sort of short exercise is often used to inform longer writing tasks. For example, the essay we used could conceivably be a step along the way (e.g., a ''problem statement'') in drafting a longer paper on the topic. In addition, the content of the writing is more generic than most discipline-based writing activities. Nevertheless, in the interest of finding a writing sample of potential relevance across disciplines, and one that would not overly tax respondents, this transportation essay provided a reasonable choice.

The composition was minimally corrected for spelling and punctuation, as those error types have not been shown to substantially affect native reader evaluation relative to other error types (Vann, Meyer, & Frederick, 1984; Sheorey, 1986). Once the composition was corrected for spelling and punctuation, the only further manipulation was to alter the first sentence of the composition so that it appeared in two guises, one Spanish and the other Chinese. To accomplish this, the first line of the composition was edited to read, ''the transportation system in [name of a city] is big problem.'' The city names inserted in the text were chosen to reflect the purported language background of the writer. Little known, neutral place names were used so as not to risk eliciting stereotypes based on the city name alone. Essentially, the ''Spanish-authored'' and ''Chinese-authored'' compositions were identical except for the place name in the first sentence. The composition appears in its entirety in Appendix B.

Evaluation procedure

Attached to each disguised composition was a brief cover letter (Appendix A) that explained the importance of learning about faculty response to international student writing. The letter twice mentioned the language group under study (L1 Chinese or Spanish writers), depending on the disguised composition attached to it. This package of cover letter and composition was then

randomly distributed to 12 departments on the campus of a large, comprehensive university in the United States. Roughly half of the surveyed departments were social science departments and the other half were math/engineering. Ten of the surveyed departments had the highest enrollments of international students on the campus.

The faculty members were asked to evaluate the composition on a 10-point scale (where 10 was supposed to represent ''excellence'' for non-native writing). The purpose of the 10-point scale was to capture a holistic rating of the composition – comparable to the scores generally given to reflective and other short writing tasks – that could be compared quantitatively across raters and across ethnic guises. Holistic scoring has been shown to be a valid method for evaluating the texts of L2 writers on the basis of author message, despite possible complications from lexical and grammatical errors compared with the texts of equivalent L1 writers (Janopoulos, 1989). While any number of scales could have been chosen, we chose the 10-point scale as it provided a wide range of scores within the standard scope of grading scales (i.e., 0–100%). Since instructions clearly stated that a score of 10 was ''excellent'' for a ''non-native'' composition, faculty were aware that they could give a top score for the essay even in terms of a non-native composition.

Faculty were also asked to mark the five most ''troublesome'' errors in the composition, which was defined for them in the instructions preceding the essay (Appendix B) as either ''irritating'' or disruptive to their comprehension. These two dimensions were intended to capture both affective and cognitive aspects of the evaluation process. Although we would be unable to tease these apart in the analysis, we felt this represented a naturalistic approach to grading an essay. Our goal was to ascertain what was troublesome to the faculty, not to establish for them ahead of time a particular type of error on which to focus. Faculty members were instructed that they could mark any kind of error (''grammatical, lexical, or otherwise'') and that they should edit the passage; that is, they should correct the errors that they encountered. The editing task was designed to probe the areas of concern to faculty. Having them not only mark the error but also correct it was important for determining which aspect of the error was bothersome to them. For example, if the word ''principle'' is changed to the spelling *principal* this could mean something different from ''principle'' corrected to *primary*. In addition, by presenting faculty with an intact composition, they would also be able to note any other types of non-sentence level errors of organization or content development.

For quantitative analysis, *t*-tests were used for comparison of means (on holistic scores) and ANOVA was used in multivariate analysis.

Results

In this section, the ''sample'' refers to the complete set of questionnaires returned, and ''sub-sample'' refers to the set of questionnaires returned for one or the other of the guises (i.e., Chinese sub-sample or Spanish sub-sample).

Study population

Of the 250 distributed questionnaires, 71 were returned (28.4%); 40 returned questionnaires were in the Chinese guise and 31 in the Spanish guise. A chi-square test indicates that this difference in sub-sample size was not significantly different ($\chi^2 = 1.6$, $p = .21$, df = 1).

Sixty-five of the 71 returned questionnaires were usable. Although this response rate is small by generally accepted standards (Baruch, 1999; Moser & Kalton, 1979), it is not statistically

relevant to analyses of the findings. We will discuss this aspect in more detail below. To be "usable," the questionnaire (Appendix B) had to have been completed with the following information:

(1) Global score rating ranging from 1 to 10.
(2) The native language of the informant, if other than English, and the number of years in the United States.
(3) The sex and the age of the informant.

Of the 65 respondents, 57 were native English speakers/writers and 8 were non-native; only 7 (all native) were female. This is roughly proportional to the larger population in the original sample (32 females, 218 males). The preponderance of male response in the survey was expected, considering that traditionally male academic disciplines were sampled for this study. To include more females in this type of research, one has to go outside of the disciplines where foreign students are enrolled in greatest number (Bellas, 1994, p. 813; Haddal, 2007, p. 12). The age range of the sample population was 29–78 years, with a mean age of 46 years.

The effect of guise and evaluator characteristics on holistic scores

Overall, the mean scores given by faculty for the compositions were quite similar regardless of the respondent's native language and regardless of the identity assigned to the composition being evaluated. Table 2 presents summary statistics for the global score ratings for the sample and sub-samples for all respondents.

The null hypothesis – that there would be no difference between the Chinese and Spanish guise scores – was not disproved in a comparison of means using an independent samples t-test ($t(50) = 0.77$, $p > .05$). Overall, there was little variability in the scoring, indicating that the global score tapped some basic holistic response to the writing sample.

Native ($n = 57$) and non-native ($n = 8$) evaluators scored the essay similarly ($M = 4.03$ and 3.75, respectively); the difference in these means, assuming non-equal variances due to the distinctly different sample sizes and the small size of the non-native evaluator sample, was not significant ($t(8.59) = .340$, $p > .05$).

The mean score from the female respondents ($n = 7$) was substantially higher (6.29) than the mean from the male respondents ($n = 58$; $M = 3.72$). A t-test, again assuming non-equal variances due to the distinctly different sample sizes, indicates that this difference in means was statistically significant ($t(6.92) = 2.85$, $p < .05$).

Holistic scores were also analyzed according to the academic departments of the respondents, social sciences (20 respondents) and engineering/math sciences (45 respondents). Table 3 presents a summary of the scores by these department groups. The means are significantly different overall between academic department categories ($t(13) = 4.19$, $p < .01$). Comparisons

Table 2
Summary statistics by guise (scale 1 → 10; 10 is high)

Guise	Mean	(S.D.)
Chinese ($n = 36$)	4.16	(2.32)
Spanish ($n = 29$)	3.79	(1.52)
Total ($n = 65$)	4.00	(1.93)

Table 3
Scores by department group

	Chinese (*n*/mean)	Spanish (*n*/mean)	Total[*] (*n*/mean (S.D.))
Social Science	10/5.55	10/4.33	20/4.9 (2.4)
English/Mathematics	26/3.65	19/3.52	45/3.6 (1.6)

[*] Only the means in the "Total" column are significantly different at $p < .05$.

between departments could not be made due to the small sample size from each individual department.

Finally, all respondents were categorized into age groups of less than 45 and greater than or equal to 45. Age, however, appeared to have no effect on the global scores. In fact, the means for the age categories were identical (4.0).

Analysis of the text-editing exercise

In addition to the research questions regarding the effect on evaluation of non-linguistic variables (implied ethnic/cultural identity of the writer as well as evaluator characteristics), the current study also set out to investigate linguistic factors affecting the evaluation of a naturally occurring L2 composition. Previous studies of error evaluation have tried to establish hierarchies of error types, often coming to conflicting conclusions about which error types belong where on the hierarchy. This is not surprising, considering the wide array of tasks and the different performance contexts (written vs. oral) in which these studies have been carried out. As mentioned at the outset, one purpose of the study was to collect, as naturally as possible, faculty evaluations of ESL writing by having respondents actually edit the passage they were confronted with, not simply respond as to whether certain formal errors seemed natural, comprehensible, or otherwise acceptable to them. The editing task attempted to elicit what faculty might actually do as teachers responsible for reading and evaluating student writing.[3]

About 50% (*n* = 32) of the respondents performed the correction task as requested by circling and correcting what they considered the 5 "most troublesome" errors in the composition. These 32 questionnaires, which were distributed nearly evenly across the two sub-samples, will be discussed first. Those respondents who circled more than 5 errors or who adopted a "rewrite style" in editing the composition will be discussed second.

Each error type discussed in this section is defined and exemplified below. The categories are based on the corrections made by the respondents, and were not established a priori. The problematic elements in the examples are printed in boldface.

Error types

ART. An article error; either omission, incorrect insertion, or incorrect article. *Example* (Line 1): *The transportation system in* [*Kao Hsiung/San Miguel*] *is* **big** *problem.*
LEX. A lexical choice error; either wrong word choice or incorrect use of a noun in the context. This also includes neologisms which are near English equivalents, but which are

---

[3] Khalil (1985) also used an editing approach in a study of response to Arabic speaking EFL learners, but that study had restricted written stimuli couched in a minimal three-sentence context.

Table 4
Error correction patterns

	ART-4	LEX-7	AUX-1	VERB-1	OMIT-2
Potential	128	160	32	32	64
% Corrected	14	33	53	38	14

ART, article error; LEX, word choice error; AUX, wrong auxiliary; VERB, tense error; OMIT, lexical omission.

clearly non-native. *Example* (Line 7): *In addition, it is* **convenient** *for the environment conservation.*

AUX. A wrong choice or use of auxiliary in the semantic sense. The tense and aspect are correct but the wrong auxiliary has been chosen for the meaning intended. *Example* (Line 10): *You* **can't** *have to wait a long time.*

VERB. A wrong verb tense, aspect, or modal is used. *Example* (Line 17): *Maybe it will change when the city* **should** *have an efficient system...*

OMIT. Lexical omission; a word is simply left out. *Example* (Line 17): *...people want to buy their own car in order to get to their workplace or other activities because the public is not efficient.*

Table 4 shows a percentage of correction for each error type according to the frequency with which the type occurred in the composition. The table is constructed to roughly compare error types and frequencies. For example, in the essay there were 4 instances of article errors (called ART-4). Given 32 respondents, there were 128 potential tokens of article correction if each respondent marked all four article errors. The second line of the table shows the actual percentage of potential tokens that were corrected.

From this table, we can see that the verbal errors (VERB + AUX) were the most often marked, as well as lexical errors (LEX). This indicates that the errors that impinge on semantic content tend to be corrected more than errors which are simply syntactic.

As mentioned, the group of questionnaires just discussed was returned by respondents who performed the editing task as directed by marking and correcting only five errors in the passage. Those respondents who marked more than five errors or who did not mark errors at all, but simply made comments about the content or organization of the composition, will now be discussed. This second group ventures into editing the non-sentence level aspects of the essay. Concerns raised by this latter group highlight faculty sensitivity to weak content or to poor organization (Full comments are reproduced in Appendix C.) Note that NS is ''native speaker,'' and NNS is ''non-native speaker:''

Respondent #8, a 29-year-old male math professor (NS): ''overall, even in proper English, the piece is poorly constructed.''

Respondent #50, a 46-year-old male civil engineering professor (NS): ''my primary concern in student writing is organization and logic. ...[T]he obvious flaws in choice of words can be easily corrected.''

Respondent #7, a 53-year-old female history professor (NNS): ''Problem is weakness of content, not of style.''

Other comments indicate that this writing sample hit another sensitive nerve for professors, an issue that appears to be more related to their experience with students who are native writers. These comments generally came from older professors, presumably those who have had the longest experience in teaching both NSs and NNSs. All of the following comments are from male professors in electrical engineering, the department with the highest overall response ($n = 11$) to the questionnaire:

> Respondent #31, a 61-year-old NS: "I give up, I'm glad this student isn't writing a thesis for me! Don't overlook the U.S. students, they need lots of English help also!"

> Respondent #32, a 57-year-old NS: "This text does not compare that unfavorably with essays I see from native-born Americans."

> Respondent #42, a 62-year-old NS: "While this essay is a disaster, it is no worse than many of the short paragraphs that my ECE [Electrical and Computer Engineering] students write."

Here the evaluative dimension ("irritation") seems clear; moreover, sensitivity to weak content and organization is not reserved for non-native students. Likewise, the reasons for such sensitivity, whatever they may be, will vary from one reader to the next. Lexicon (cf., Engber, 1995) and grammar may disrupt one reader's response, whereas another is more concerned that, for example, an idea mentioned in an introductory paragraph is never addressed in the body of the essay (Respondent #47, a 52-year-old male sociology professor). In addition, it is evident that some faculty simply will not accept a text as it is written, regardless of the correctness of the language and content. These professors are engaged in a "rewrite" style of editing that may have little to do with the composition or its writer and more to do with the teacher's preference for certain vocabulary or for a particular rhetorical style. For example, one professor changed every instance of the word "car" to *automobile* and changed "talk about" to *discuss* (Respondent #42, described above). Another professor changed "is not efficient" to *is inefficient* (Respondent #10, 34-year-old male NNS in chemical engineering). With this kind of rewriting editor/teacher, one may never have a correct or acceptable prose style, regardless of native language.

Discussion and conclusion

This study used a matched-guise technique to explore the role of writer characteristics in the evaluation of ESL writing by university faculty. The study also examined the types of corrections made by these professors, attempting to characterize the sorts of errors that they find most disruptive. Although we did not distinguish affective from cognitive responses to a writing sample, we aimed to provide a more naturalistic, holistic task for study participants to perform. By using a relatively short writing passage, one that was identifiable as academic and not personal, we were able to impinge less on faculty time while still presenting a topic that would be relevant to a wide variety of academic disciplines. In this section, we first discuss the findings concerning writer identity, then the error correction task, and finally the findings concerning evaluator characteristics.

The holistic ratings and error correction patterns were not influenced by the presentation of implied ethnic/cultural identity for the writer. In contrast to Rubin and Williams-James (1997), which examined the effect of non-native speaker/writer identity against native writer identity and

found important qualitative and quantitative differences in reader response (to Thai, Danish, and American guises), the current study found no effect for writer identity. Apparently the negative stereotype sometimes associated with Spanish-accented speech (Brown et al., 1985; Ryan & Carranza, 1975) and with other ESL learners' performance (Delamere, 1986) is not associated with a writing sample presented as though from an L1 Spanish writer. Our expectations of a transfer of negative stereotypes (e.g., Fishman et al., 1987) or even evaluator disposition to counter stereotypes by softening or reducing criticism (Rubin & Williams-James, 1997) were not supported. The fact that neither the ''Spanish'' nor the ''Chinese'' guise negatively affected the scoring of the student composition requires further consideration of whether possible negative stereotypes associated with an ethnic group are transferable to written media and to what degree. At present, it seems that these university professors were not negatively influenced by students' ethnic background. We conclude that when ESL linguistic features are held constant, the faculty did not rate the student composition based on ethnic guise.

However, the results of this study corroborated findings from earlier studies of the effect of academic discipline on the evaluation of ESL compositions (Santos, 1988; Vann et al., 1984). To date, much of the research done on population factors and response has focused on peer review (cf., Leki, 2001) rather than the higher-stakes circumstances of instructor response. The current study underscores the important relationship between professors' academic disciplines and their different orientations to the evaluation of written work. This finding indicates the need to explore in further detail the attitudes and expectations of university professors across disciplinary boundaries.

The response to the errors in the sample composition indicated that unclear meaning traceable to poor word choice and to verb tense or modal errors were focal points of correction by the faculty. The error correction frequencies were consistent with observations from previous studies, namely that native readers are most sensitive to verb errors (Piazza, 1980; Politzer, 1978; Sheorey, 1986) and lexical errors (Khalil, 1985; Sheorey, 1986). Granted, ''sensitivity'' has been measured in different ways by different researchers, but the outcome has been fairly consistent across languages, despite the variability in methodology. The findings of this study indicate that even when given a natural passage to edit, evaluators still choose verbal (i.e., tense/aspect) and lexical errors to correct, essentially those errors which ''impinge directly on content'' (Santos, 1988, p. 84). Overall, it seems that a semantic gap (whether wrong word choice or wrong tense/aspect) is more frequently edited than grammatical items such as articles, indicating a preference by the faculty to clarify content, a finding that is supported by conclusions from Santos (1988, p. 80).

Importantly – but not something that composition instructors may be able to do much about some faculty respondents simply rewrote the essay regardless of its technical correctness. Apparently, they wanted the composition written with certain vocabulary or in a particular style that they preferred. Some faculty members were concerned about grammar/lexicon, while others did not mark any sentence-level errors. This division may simply be an artifact of who followed questionnaire directions and who did not, but it is likely more an echo of results first noted by Engber (1995, pp. 149–150), in which overall lexical variation and lexical variation minus error were highly correlated with holistic scores. Variation in this case may not be an intra-textual matter, but also an *inter*-textual one, compared with the more generic language of broader English-using society. Perhaps English for Academic Purposes (EAP) instructors might prepare students for the variability of response to their writing by having, not only peer review, but also cross-teacher review. Occasionally, teachers may want to exchange their students' writing assignments to provide the students with additional and different editing input that can better prepare them for the variability of response they are likely to experience.

Age, when treated categorically (<45 or >45), showed no effect on the global scores. This finding differs from studies of similar university faculty populations (Santos, 1988; Vann et al., 1984) which showed more grading leniency among older professors.

Although our sample of female respondents was small, the mean score for that group was significantly higher than the mean score for the male respondents. This finding is in line with a study of L1 written peer reviews which demonstrated that males and females differed in their patterns of complimenting and personal referencing in written discourse (Johnson & Roen, 1992), indicating not only a difference in discourse style, but also perhaps an overall gender difference in response to writing tasks. In prior studies, Santos (1988) did not report any specific findings regarding the female sample in her study ($n = 22$), and in Vann et al. (1984), analysis of the effect of gender on evaluations of L2 writing had to be "thrown out" due to small sample size. We do not believe that the finding of gender difference in this study provides a practical point of departure for further exploration, but it does highlight the variability of evaluator response in terms of personal characteristics. We choose to report this finding since earlier studies have addressed the possibility of a gendered response to the evaluation process.

In addition to the small size of the female sample, the response rate for the study was below a generally accepted threshold of 40% for mailed-out surveys (Baruch, 1999; Moser & Kalton, 1979), but this threshold does not have statistical relevance for testing findings. The main concern with a low response rate is determining whether or not non-responders are substantially different from those who did respond. In the current study, we can assume a fairly homogenous target population in terms of socio-economic characteristics; all of those invited to participate were university professors. What we have lost by the low response rate is some level of confidence in our ability to capture and account for the widest range of variability in professors' responses to the ESL writing; otherwise, the sample size itself is sufficient for the standard inferential statistics used for this study.

In textual terms, the essays regularly assigned as major projects in composition are generic in nature, meant to reflect the most common elements of academic writing, rather than the specifics of genres in the various disciplines or the kinds of discipline-specific responses associated with them. However, without departing from introductory EAP's general, cross-discipline language-support role, instructors might find opportunities in their curricula to briefly examine work from the major area of each student. Such activities might be comparable to the approach described by Crompton (2005), in which students compared their own work to that of professional writers. Providing EAP students with instructor feedback on discipline-specific, student-written texts in addition to professionally written samples might provide the kind of practical exercise that would further build L2 student awareness of connections between features of writing and instructor response.

In summary, these findings should encourage EAP instructors on university campuses to devote greater attention to variability in response for students who will be continuing to engineering and math sciences, as well as for those in other fields. It appears from this study, and from prior research, that professors in engineering and math sciences are slightly more exacting in their expectations of ESL writing. This is reflected in lower scores given to the ESL essay by faculty in math/science disciplines. Since international students tend to enroll in greatest numbers in non-humanities programs (Haddal, 2007, p. 12), it is crucial to provide the kind of writing instruction that will prepare them to enter those fields. In an informal post hoc discussion with a small portion of those faculty who volunteered to meet following the survey ($n = 3$), it was made clear to the first author that international students are not being trained to do the kind of writing that is expected of them in scientific disciplines. Although only anecdotal in nature, this

commentary from the faculty serves as a cautionary note. Writing instructors could find it beneficial to attend closely to faculty attitudes as they work to develop curricula and activities that will be relevant for students as they transition out of EAP instruction.

References

Aigner, D., & Thum, F. (1986). On student evaluation of teaching ability. *Journal of Economic Education, 17*, 243–265.

Baruch, Y. (1999). Response rate in academic studies: A comparative analysis. *Human Relations, 52*, 421–438.

Beason, L. (2001). Ethos and error: How business people react to errors. *College Composition and Communication, 53*(1), 33–64.

Bellas, M. L. (1994). Comparable worth in academia: The effects on faculty salaries of the sex composition and labor-market conditions of academic disciplines. *American Sociological Review, 59*(6), 807–821.

Bradac, J. J. (1990). Language attitudes and impression formation. In H. Giles & W. P. Robinson (Eds.), *The handbook of language and social psychology* (pp. 387–412). John Wiley & Sons.

Bradac, J. J., Cargile, A. C., & Hallett, J. S. (2001). Language attitudes: Retrospect, conspect, and prospect. In W. P. Robinson & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 137–155). John Wiley & Sons.

Brown, B. L., Giles, H., & Thakerer, J. N. (1985). Speaker evaluations as a function of speech rate, accent and context. *Language & Communication, 5*(3), 207–220.

Callan, V., & Gallois, C. (1983). Ethnic stereotypes: Australian and Southern European youth. *The Journal of Social Psychology, 119*, 287–298.

Chastain, K. (1980). Native speaker reaction to instructor-identified student second language errors. *Modern Language Journal, 64*, 210–215.

Clayton, M. (2000). Foreign teaching assistants' first test: The accent. *Christian Science Monitor, 92*(198), 14.

Crompton, P. (2005). ''Where,'' ''In which,'' and ''In that'': A corpus-based approach to error analysis. *RELC Journal, 36*(2), 157–176.

Delamere, T. (1986). *The role of stereotyping in native speaker judgments of English as a second language learners' performance*. Unpublished Doctoral Dissertation, Florida State University.

Delisle, H. H. (1982). Native speaker judgment and the evaluation of errors in German. *Modern Language Journal, 66*, 39–48.

Eisenstein, M. (1983). Native reactions to non-native speech: A review of empirical research. *SSLA, 5*(2), 160–176.

Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139–155.

Fishman, G., Rattner, A., & Weiman, G. (1987). The effect of ethnicity on crime attribution. *Criminology, 25*, 507–524.

Guntermann, G. (1978). A study of the frequency and communicative effects of errors in Spanish. *The Modern Language Journal, 62*, 249–253.

Gynan, S. N. (1985). Comprehension, irritation and error hierarchies. *Hispania, 68*, 160–165.

Haddal, C. C. (2007). *Foreign students in the United States: Policies and legislation*. Washington, DC: Congressional Research Service.

Hairston, M. (1981). Not all errors are created equal: Nonacademic readers in the professions respond to lapses in usage. *College English, 43*(8), 794–806.

Harklau, L., Losey, K. M., & Siegal, M. (Eds.). (1999). *Generation 1.5 meets college composition* (pp. 191–210). Mahwah, NJ: Lawrence Erlbaum Associates.

Janopoulos, M. (1989). Reader comprehension and holistic assessment of second language writing proficiency. *Written Communication, 6*(2), 218–237.

Janopoulos, M. (1992). University faculty tolerance of NS and NNS writing errors: A comparison. *Journal of Second Language Writing, 1*(2), 109–121.

Johansson, F. (1978). *Studies of error gravity: Native reactions to errors produced by Swedish learners of English*. Gothenburg Studies in English 44. Gothenburg, Sweden: Acta Universitatis Gothoburgensis.

Johnson, D. M., & Roen, D. H. (1992). Complimenting and involvement in peer reviews: Gender variation. *Language in Society, 21*(1), 27–57.

Khalil, A. (1985). Communicative error evaluation: Native speakers' evaluation and interpretation of written errors of Arab EFL learners. *TESOL Quarterly, 19*(2), 335–351.

Lambert, W. E., Hodgson, R., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal & Social Psychology, 60*, 44–51.

Leki, I. (1995). Coping strategies of ESL students in writing tasks across the curriculum. *TESOL Quarterly, 29*(2), 235–260.

Leki, I. (2001). "A narrow thinking system": Nonnative-English-speaking students in group projects across the curriculum. *TESOL Quarterly, 35*(1), 39–67.

Leonard, D. J., & Gilsdorf, J. W. (1990). Language in change: Academics' and executives perceptions of usage errors. *Journal of Business Communication, 27*(1), 137–158.

Ludwig, J. (1982). Native-speaker judgments of second-language learners' efforts at communication: A review. *Modern Language Journal, 66*, 274–283.

Magnan, S. S. (1983). Age and sensitivity to gender in French. *SSLA, 5*(2), 194–212.

Moser, C. A., & Kalton, G. (1979). *Survey methods in sociological investigation*. London: Heinemann.

Norris, T. (1991). Nonnative English-speaking teaching assistants and student performance. *Research in Higher Education, 32*(8), 385–397.

Piazza, L. (1980). French tolerance for grammatical errors made by Americans. *The Modern Language Journal, 64*, 422–427.

Politzer, R. L. (1978). Errors of English speakers of German as perceived and evaluated by German natives. *Modern Language Journal, 62*, 253–261.

Ranelli, P. L., & Nelson, J. V. (1998). Assessing writing perceptions and practices of pharmacy students. *American Journal of Pharmaceutical Education, 62*(4), 426–432.

Rifkin, B., & Roberts, F. D. (1995). Error gravity: A critical review of research design. *Language Learning, 45*(3), 511–537.

Rubin, D. L., & Williams-James, M. (1997). The impact of writer nationality on mainstream teacher's judgments of composition quality. *Journal of Second Language Writing, 6*(2), 139–153.

Ryan, E. B., & Carranza, M. A. (1975). Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. *Journal of Personality and Social Psychology, 31*, 855–863.

Santos, T. (1987). Markedness theory and error evaluation: An experimental study. *Applied Linguistics, 8*(3), 207–218.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly, 22*(1), 69–90.

Sheorey, R. (1986). Error perceptions of native-speaking and non-native-speaking teachers of ESL. *ELT Journal, 40*(4), 306–312.

Silva, T. (1993). Toward an understanding of the distinct nature of second language writing: The ESL research and its implications. *TESOL Quarterly, 27*(4), 657–677.

Tomiyana, M. (1980). Grammatical errors communication breakdown. *TESOL Quarterly, 14*(1), 71–79.

Vann, R. J., Meyer, D. E., & Frederick, L. O. (1984). Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly, 18*(3), 427–440.

Williams, J. M. (1981). The phenomenology of error. *College Composition and Communication, 32*, 152–168.

Zamel, V. (1995). Strangers in academia: The experiences of faculty and ESL students across the curriculum. *College Composition and Communication, 46*(4), 506–521.

Appendix A. Cover letter

Date

Dear Faculty Member:

I am writing to ask for your assistance in an important research project concerning the English writing skills of international students whose first language is [Chinese/Spanish]. As a teacher of English as a Second Language on the university campus, I am concerned about training students to write for an academic audience. I believe that part of that training should involve input from faculty members regarding their response to non-native English writing. To that end, I am conducting research this semester to ascertain faculty response to English writing samples from native [Chinese/Spanish]-speaking students. Please take a moment to read the attached composition and questionnaire which can be filled out in a few minutes and returned to me via campus mail. Your participation is completely voluntary, and your responses will never be identified with you.

Thank you in advance for your assistance with this project.

Sincerely,

Researcher name

Department of English, Program in ESL

Appendix B. Composition and questionnaire

B.1. Directions

Please read the following composition (it continues onto the back of this page.) After reading it, you will find a short series of questions to answer. First you will be asked to rate the composition on a scale of 1 to 10, where 10 represents excellence for a non-native composition. Then, once you have given a global rating to the writing sample, please circle and correct the 5 most troublesome errors for you as a reader. By ''troublesome'' I mean those errors which interfere with your reading, either because you find them irritating or because they disrupt your comprehension of the passage. Whatever the error, whether it is grammatical, lexical, or otherwise, please indicate the correction you would make as if you were editing the passage. Finally, please answer the short series of demographic questions. Thank you again for your time and attention.

Please return by [date] via campus mail to: Researcher [address].

B.2. Urban transportation system, almost always a big problem

The transportation system in [Kao Hsiung/San Miguel] is big problem. Some people like taking a public transportation and others prefer driving their own car. Mostly people think about saving money and safety when they talk about urban transportation.

Public transportation is not efficient because the citizens have to wait for a long time and it requires at least one or two transfers to get where you are going. But the bus system is more economical service because it is a massive system of transportation and the passengers don't have to look for a parking place like with a particular car. In addition, it is convenient for the environment conservation because the cars are a principle source of pollution.

On the other hand, driving a particular car is convenient because you can get to any place you want to go; moreover, you can't have to wait a long time. However, there are some disadvantages such as a parking place, the cost of maintenance, and traffic problems. Also, it has an environmental implicases because each car is a pollutant factor.

About the public transportation in my city, it will be important to plan a better system like a subway. Nobody is concerned about this kind of service, but it will be more efficient and valuable for the environmental conservation. Nowadays mostly people want to buy their own car in order to get to their workplace or other activities because the public is not efficient. Maybe it will change when the city should have an efficient system of massive transportation. It is a big task for the policy makers.

♦   ♦   ♦   ♦   ♦   ♦   ♦   ♦   ♦   ♦   ♦   ♦

PLEASE COMPLETE THE FOLLOWING QUESTIONS:

1.   Rating for the composition:
     (Circle one number: 1 is low, 10 is high)

     1   2   3   4   5   6   7   8   9   10

2.   Indicate the 5 "error spots" and corrections on the composition itself.

3.   Demographic Information:

     ♦ Your native language: _____ (If other than English, how many years
     have you been in the U.S. or other country where English is used for instruction?
     Please note the total number of years _____)
     ♦ Sex: FemaleMale
     ♦ Age:
     ♦ Number of international students you are teaching and/or advising this
     semester:_____

4.   I am interested in getting faculty input regarding English language training of
     international students on this campus. Would you be willing to participate in a discussion
     among faculty concerning the training of international students for academic writing?

     Yes (please indicate your name and phone number):

     No

Appendix C. Respondent comments

C.1. Irritation/comparison with Americans

(1) Too many [errors] *all* are a problem. We find that most require considerable additional
    training. I have no (zero) tolerance for errors in papers, theses, etc. which are to be
    published—so often considerable editing is required by the faculty [30].
(2) I give up, I'm glad this student isn't writing a thesis for me! Don't overlook the U.S. students,
    they need lots of English help also! [31]
(3) It's hard to choose 5 [errors] from this very poorly written sample [36].
(4) While this essay is a disaster, it is no worse than many of the short paragraphs that my ECE
    students write [42]
(5) . . .30 or so stylistic slips – I only marked 5 as you requested. This text does not compare that
    unfavorably with essays I see from native-born Americans – perhaps you could broaden your
    project? [32]
(6) Most of the problems with this essay have little to do with grammar. (Sometimes it is
    impossible to decide what the author meant, so it's not clear what correction is indicated.)
    For example, the entire point could have been gotten across in two or three sentences! Except
    for grammar/spelling problems, this is not much worse than essays written by U.S. students!
    [22]

(7) [25] ''Bad writing!''; [5] ''I gave up after the first paragraph''; [45] edits first few lines, then writes, ''etc.'' and leaves the rest blank; [10] ugh!

C.2. Organization/content

(8) My primary concern in student writing is organization and logic [needs to know purpose to evaluate effectiveness]... as a technical article, it is poorly organized and excessively wordy. The obvious flaws in choice of words can be easily corrected [50].

(9) Problem is weakness of content, not of style [comment on 10 rating—[7].

(10) Overall, even in proper English, the piece is poorly constructed [8].

C.3. Language background

(11) I thought this was very good for a Chinese writer. I wonder what this person's spoken English is like [9].

(12) In our Dept. (Chem. E) the Asian students usually have weaker language skills than Hispanics [58].

C.4. Grammar/lexicon

(13) [Comment on a ''5'' rating) but I rarely get something this clear. Main problem here is articles. Main problem my students have is using abstract words with no substance behind them [17].

C.5. Other

(14) I am more concerned with skills in listening and speaking since the writing, while discordant, is normally intelligible [28].

Felicia Roberts is an associate professor in the Department of Communication at Purdue University. Her research focuses on the coordinated verbal and non-verbal practices that are assembled in the construction of identities, and that affect perceptions, in everyday and institutional life.

Tony Cimasko is currently a doctoral candidate in the English Department of Purdue University. His research focuses on the contextual dimensions of academic and professional genre texts of both first and second language writers.