

Working Paper

Assessment Across Borders: National Perspectives Explain Differences Between Singaporean and US Evaluators

Colin M. Clark

Nanyang Business School
Nanyang Technological University, Singapore

Priscilla S. Rogers

Stephen M. Ross School of Business
at the University of Michigan

Ross School of Business Working Paper Series

Working Paper No. 901

November 2004

This paper can be downloaded without charge from the
Social Sciences Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=901653>

**Assessment across Borders: National Perspectives Explain
Differences between Singaporean and US Evaluators**

Colin M. Clark

Nanyang Business School, Nanyang Technological University, Singapore
S3-B1c-104, 50 Nanyang Ave. 639798
Singapore
Phone: +65 6790-6034 Fax: +65 6792 4217
Email: acmclark@ntu.edu.sg

Priscilla S. Rogers

University of Michigan Business School
701 Tappan Street,
Ann Arbor, MI 48109-1234
USA

Phone: +1 734 764 9779 Fax: +1 734 936 0279
Email: psr@umich.edu.sg

Abstract

This study suggests that analytical tools to assess writing across genre can be meaningfully used across different countries. However, evaluators' "national perspectives" are likely to impact the assessment of content, particularly as it relates to completing the writing task. We compared Singaporean and American evaluators' assessment of written responses to workplace scenarios, requiring critiquing a superior's ideas. Two responses were collected from upper-level business school students at a major university in the Republic of Singapore: one response prior to and another at the end of a business communication course. Holistic scores of this corpus were used as a basis for selection of a core sample of 468 responses, which Singaporean and US evaluators independently scored on four analytical tools: task, reasoning units, coherence, and error interference. US evaluators gave significantly higher scores on task fulfillment and reasoning units than did Singaporean evaluators, and only the US evaluators found improvement in the post-assessment compared with the pre-assessment. Subsequent textual analyses suggested that these differences stemmed from content preferences we characterize as national perspectives--US evaluators favored an external "proactive" focus based on potential gains, whereas Singaporeans preferred an internal focus based on avoidance of potential losses. This finding has implications for cross-national education, assessment and training.

Assessment Across Borders: National Perspectives Explain Differences Between Singaporean and US Evaluators

As the workplace becomes increasingly global, there are signs of renewed interest in accountability, standardization, and in transferable vehicles for testing, certification, and benchmarking student performance across borders, including in business communication programs preparing new hires. For example, prestigious schools throughout the European Union have formed the Consortium of European Management Schools (CEMS) that facilitates a joint assessment program to measure student language preparation for group interaction. Meanwhile, assessment vendors such as Educational Testing Service continue to develop new instruments and refine old, supporting research for such efforts including the TOEFL monograph series, phase one of which examines test evaluators themselves (Cumming, Kantor & Powers, 2001).

Leading the way in Asia, Singapore recently implemented the Scholastic Aptitude Test while other assessments for university admissions and placement, such as the English Qualifying Examination and the Graduate Management Admissions Test (GMAT) with its Analytical Writing Assessment are now entrenched there. As with the CEMS collaboration in Europe, the largest business school in Singapore joined some US business schools in offering case-based performance assessments to provide feedback and to quantify student improvement. In conjunction with this, Singaporean and US faculty began a conversation that reawakened questions about whether communication effectiveness can be evaluated consistently across borders.

If it is important to evaluate the preparedness of students for the global workplace, then it is desirable that evaluators around the world share judgments of

quality, is it not? Evaluators of such tests as the International English Language Testing System (IELTS) or GMAT are given descriptors and training that do not vary between countries, on the assumption that a judgment of writing quality from an evaluator in one country will be equivalent to one made by an evaluator with the same disciplinary focus in another. We question whether this assumption is justified.

We know that teams of evaluators within a country assess consistently; indeed, such has come to be expected when there is sufficient training on a scoring scheme with a set of benchmark writing examples (Cumming, Kantor & Powers, 2001). But even with rigorous training, research shows that across countries, evaluation teams apply the same scoring scheme differently (Purves, 1992). Indeed, some dismiss the goal of evaluating consistently across borders, believing that such attempts are reductive and deterministic and further that they gloss over the fact that any group is highly idiosyncratic (Kubota, 1999; Siegel, 1997; Zamel, 1997). Others suggest that rather than attempting to apply the same evaluative schemes across groups, evaluation might in some way acknowledge each group's unique cultural contexts and anticipate that certain competencies may be valued to a greater degree in some contexts than in others (Atkinson, 1997; Carson, 1998; Ramanathan & Kaplan, 1996a, 1996b).

If we agree that evaluation is an important aspect of programs intended to prepare individuals to communicate effectively in the global workplace, then we need to compare our judgments of quality to explore issues of consistency. Might consistency in evaluation across borders be achieved if evaluators shared a professional, disciplinary focus and pedagogical imperatives, such as business school faculty training soon-to-be hires for the workplace?

Purpose

This study revisits the issue of writing evaluation across countries but with a narrower focus than the well-known International Association for the Evaluation of Educational Achievement (IEA) study (described below). We focused on teams of faculty evaluators in business schools who were responsible for training upper-level university students, or soon-to-be new hires for the workplace. One team of evaluators worked at a major business school in Singapore, the other at a top-tier business school in the Midwestern US. Much like the IEA study, our teams of evaluators were trained under the very same protocol. Moreover, since the primary concern was preparing business students for the workplace, we employed scenarios as assessment prompts that required students to respond to workplace writing tasks. We evaluated their responses holistically and analytically using tools developed specifically for business education (Rogers & Rymer, 2001). Our primary intent was to explore whether evaluators who shared a disciplinary focus and many pedagogical goals, but who lived and worked in two quite different national environments would apply these targeted tools differently. As reported in our Results section, the correlative data we obtained via the holistic and analytical scoring led us to conduct a series of textual analyses of several sub-samples, including of responses where we discovered quite dramatic disagreement between evaluation teams.

Literature Review

This research addresses questions generated by prior studies on the use of analytical tools across groups, genre, and countries. It also extends what we know about textual and contextual matters that influence evaluators' decisions regarding quality.

Using Analytical Tools across Groups, Genre & Countries

By requiring evaluators to make decisions about particular features--rather than the composition as a whole as for holistic evaluation--analytical tools have long been viewed as a way to identify linguistic and rhetorical features affecting evaluators' judgments (Connor, 1990; Lloyd-Jones, 1987; Rogers, 1994). So it is not surprising that analytical tools have been developed and used to compare writing evaluation across groups, genre, and countries, with some failure and some success.

For the International Study of Written Composition, a highly ambitious ten-year study sponsored by the IEA, evaluators scored a variety of writing samples (e.g., letter of advice, narrative writing, and argumentative writing) written by thousands of children (ages 12, 16, and 18) from 14 different countries (Connor, 1990; Gorman, Purves, & Degenhart, R. E., 1988; Purves & Takala, 1982). For this massive cross-national evaluation, an analytic scoring scheme was developed to facilitate comparisons of quality, content, organization, style, lexical and grammatical features, spelling and orthographic conventions, as well as handwriting and neatness. (Purves, Gorman, & Takala, 1988). In the end, comparison across borders was not achieved, leaving Purves to conclude quite boldly that the study ended in failure (Purves, 1992).

Evaluating only the persuasive essays in the IEA study written by students from three English-speaking countries, Britain, New Zealand, and the US, Connor (1990) experienced more success. Quantifiable scales were found to be reliable, to isolate features that contributed to overall quality, and to enable detailed comparisons among linguistic and rhetorical features used by students from these three English-speaking countries. For example, compared with British and New Zealanders, US writers used less data, a less passive and nominalized style, and demonstrated a preference for colloquial and interactive features (Connor & Lauer, 1988; Connor & Biber, 1988). As

Connor concluded, these results have implications for the teaching of persuasive writing (Connor, 1990).

Targeting assessment for business education, Rogers and Rymer (2001) developed analytical tools to bridge two quite different types of writing in an attempt to provide feedback to students entering MBA programs. Their tools for evaluating task fulfillment, coherence, reasoning units, and error interference were originally designed to diagnose MBA applicants' writing skills using the essays written for the GMAT Analytical Writing Assessment as a way to help these new entrants understand how writing strengths and weakness might impact their performance on the quite different kind of writing they would be required to produce for their MBA studies. In other words, these analytical tools were intended to bridge contexts, identifying fundamental traits of writing that were significant for two quite different types of writing--the essay on the one hand and various written responses to management cases on the other. In addition to bridging genres, Rogers and Rymer's (2001) research also demonstrated that the tools could be employed across two quite different US business school contexts, a top-tier elite business school and a mainstream school, leading them to invite researchers to explore whether evaluators in two quite different cultural contexts might employ the tools meaningfully.

Taken together, this research suggests that while it may not be possible to develop instruments that can be applied consistently across many genre, age groups, and nationalities, it is possible to develop analytical tools targeting various disciplinary or professional groups. Rogers and Rymer's (2001) work suggests that evaluators in different US contexts but with the same disciplinary and professional focus share values regarding writing quality in relationship to task, coherence, reasoning, and errors. Would this hold true across countries? We know from Connor's (1990)

research that analytical tools focusing on a particular type of writing provide useful comparative data across different English-speaking countries. One question for this research was to investigate if Rogers and Rymer's (2001) analytical tools for business writing would also yield reliable and meaningful cross-country comparisons, particularly for assessment.

Influences on Evaluation

Research further shows that evaluators are influenced most by content development, including Western forms of argument that are unfamiliar to non-native speakers of English. Disciplinary expectations also play a role in differing evaluator perceptions.

As research has shown for some time, of all the writing characteristics of interest to evaluators, the nature of the content has an impact on both holistic and analytic assessment of effectiveness. For example, Harris (1977) found a tendency for teachers to give the most weight to content and organization when evaluating student essays. In a follow-up study involving a larger sample of argumentative essays and more systematic research design, however, Freedman (1977; 1979a & b) found content to be the most significant influence on evaluators' holistic scores. Using a four-point holistic scale, evaluators were influenced by content (the development and logical presentation of ideas) first, then organization (order, transitioning, and paragraphing), and to a lesser degree sentence structure and mechanics (usage and punctuation). "The difference between the average score given papers strong in content versus the average score given papers weak in content was 1.06," Freedman reported (1977, p.163). Sometime later, Breland and Jones (1984) reconfirmed this conclusion.

In a series of studies Connor and Lauer identified a number of content variables as useful for predicting the overall quality of student persuasive writing, such as the

quantity and types of persuasive appeals (Connor, 1990; Connor & Lauer, 1985, 1988) and audience awareness (Connor, 1987, 1990). Related to this they found that non-native speakers of English did not score as well as native speakers on an analytic tool based on Toulmin's requirements for the development of an argument (e.g. how to formulate an effective claim, how to use data to support that claim, how to use warrants to link the data to the claim). Building on their work, Ferris (1994) showed that native speakers were more inclined to produce counterarguments, another aspect of Toulmin's argumentation, than non-native speakers. Like Connor and Lauer, he concluded that these differences stemmed in part from the fact that non-native speakers were writing from a different discursal and rhetorical framework (perhaps attributed to differing rhetorical conventions in their native languages) than native speakers (Connor, 1990, 1987; Connor & Lauer, 1985, 1988; Ferris, 1994).

This research recalls Shepherd and O'Keefe's (1984) work on reader adaptation. They found that some types of content seemed to be more appealing to readers than other types of content. "Constructing an effective message is not a matter of generating just any message to fit some abstract pattern," they concluded, "but rather of exploiting the information available in the situation to construct the most effective specific appeal" (1984, p. 151). As readers, we might expect evaluators to be influenced in one way or another by the extent to which the writing is adapted to the particular concerns, interests, and perhaps even cultural experience of the evaluators.

In addition to differences between native and non-native speakers, an established stream of research illustrates that social contexts of writing and disciplinary experience impact expectations for writing (e.g., Bazerman, 1988; Bizzell, 1992; Herrington, 1985; McCloskey, 1985; Swales, 1998). Studies investigating writing in a wide range of disciplines, from engineering to literature, from biology to English

composition, show that different disciplines have different expectations for writing. Looking at classroom interactions, Herrington (1985) and Doheny-Farina (1986) showed that values, language, thought processes, and genre expectations were evident in discussions of content. Observing the discourse of three academic communities, each organized on a different floor in a small, three-floor university building, Swales (1998) found unique discourse particulars stemming from the specializations of each group.

Research also demonstrates that expectations for writing vary with disciplinary focus in the professional world. Written rhetoric is guided by certain workplace conventions involving issues, ideas, and arrangements (see also Brown & Herndl, 1986; Paradis, Dobrin, & Miller, 1985; Rogers, 1989; Winsor, 1996, 1999). This research suggested that the business communication focus of the prompts, evaluation tools, and faculty evaluators might influence our evaluation in some way. If evaluators shared a disciplinary focus and obligation to prepare the respondents for the workplace, perhaps we would find more agreement than resulted from the IEA study despite that fact that they crossed cultures.

In the aftermath of the IEA study, Purves (1992) stated he saw no particular need for more research on scoring, prompt development, or the merits of different scoring schemes, yet he expressed continuing curiosity about how evaluators' perceptions color their judgments. People in the arts and in sports, he concluded,

. . . trust the judgments only so far as they trust the jury. In many fields jurors go through extensive training. In writing assessment, the training tends to be about average and in writing research it tends to be minimal. No matter how extensive or thorough it may be, the rating is still a perception, a subjective estimate of quality (1992, p. 118).

For some time the notion that a text itself is autonomous and can be judged against some absolute standard or template has been rejected, of course. Meaning exists, Raymond explained, "not exclusively or even primarily in the text itself, but in the several marriages each text makes with the minds of its individual readers" (1982, p. 400; see also Gere, 1980). Given this, Raymond argues that the chief value of training evaluators may be less to achieve inter-rater reliability but more to prompt them to examine their assumptions and to arrive at agreements about what is important and unimportant in writing for their consistencies. Consensus reached at one institution, he concluded, will and ought to vary from the consensus reached at another institution (Raymond, 1982).

Speaking of the Singaporean context specifically, Chee (1996) reminds us the while English competency is high among university students in Singapore and the US and while English is the language of instruction in both countries, there are differences that may impact evaluators' perceptions of quality:

Writing is fundamentally a social process, and as such, it is influenced by the philosophical foundations and value systems of the society in which it is grounded. In Singapore the medium of instruction is English while many of the values that underlie Singaporean society and education derive from Chinese, Malay and Indian culture (1996, p. 35).

These streams of research suggest that content development and related issues of native/non-native language background, disciplinary and professional experience, and individual perception impact evaluation. Therefore, we might expect to observe some of these issues coloring the conclusions of the Singaporean and US evaluation teams in our study, even though the teams were uniformly trained, used the same analytical tools and scoring protocol, and evaluated the very same writing samples. We set out to

understand how evaluation teams from the two English-speaking countries might differ in applying analytical tools designed to assess features valued for workplace writing.

Building on this research, questions addressed in this study include the following:

- Would analytical scoring on task, coherence, reasoning and error interference reveal improvement not captured by holistic evaluation?
- Would consistency in evaluation across borders be achieved if evaluators shared a professional, disciplinary focus, and pedagogical imperatives?
- Would the disciplinary focus of the analytical tools (designed for business training) and of our evaluation teams (business communication faculty) impact inter-evaluator team agreement?
- Would Singaporean and US evaluator judgments differ on analytical tools most closely tied to content issues (the task fulfillment and reasoning units tools) and less so tools tied to organization (coherence tool), sentence structure, and mechanics (error interference tool) as suggested via previous research?
- Would contextual differences between the Singaporean and US evaluation teams impact the evaluation in any way?

Background on the Study

This research began as an exploratory study in the largest business school in Singapore, a study that produced puzzling results and prompted the more ambitious project we report here. Using a performance assessment methodology, our original goal was to collect qualitative and quantitative data on upper-level undergraduates' writing skills both before and after they had taken a business communication course in order to identify needs and measure learning. Our pre- and post-assessment prompts (piloted) were workplace scenarios, comparable writing tasks involving a request from

a boss to critique his argumentation. Students' responses to these tasks were scored holistically by teams of faculty using the six-point scoring scale following an ETS-based protocol (holistic scoring criteria and protocol available upon request). In total, 1226 responses were collected and scored holistically, including pre- and post-assessment responses from 613 individual students. The intent was to mine these data to provide feedback to students, faculty, and administrators. In fact we were quite sure that this exercise would allow us to quantify the success of the business communication training.

Surprisingly, while the holistic scores of some students appeared to have improved for the post-assessment, those of many others had declined. More perplexing was the fact that the scores showed no significant ($\alpha = .05$) change between the pre- and post-assessment. This finding suggested a number of different explanations. Perhaps this result was due to a weakness in the scoring methods or protocol? This seemed less likely than other explanations, however, as the holistic scoring procedures involved established assessment controls rigorously applied including evaluator training, blind scoring of each response by at least two evaluators, group blind readings of every 35th response, and monitoring by table leaders to insure a high degree of inter-rater reliability.

Of more concern were the following questions: (1) Was the result an accurate reflection that the students had not progressed, or was it a mismatch between evaluator expectations and the communication strategies that the students were applying? (2) Were any of the assessment criteria elevated above the others in the formulation of our evaluator's judgments, skewing the evaluation in some way? (3) When some students had substantial increases of 2-3 scoring levels between the pre- and post-assessment, why had so many students' scores decreased by the same amount? (4) Would

evaluators removed from the local context make the same judgments of quality that our Singaporean evaluators did? (5) What implications did the findings have for our efforts as an education provider in the global marketplace?

To address these questions we formed two new evaluation teams--one at the Singaporean site of the exploratory study, the other at a comparable US business school. We formed these teams to analytically score a sample drawn from the original corpus (described below) using a set of tools designed for the business schools (Rogers & Rymer, 2001). If these teams of evaluators again failed to find significant differences between pre- and post-assessments using analytical tools, we would have validation for the original finding based on holistic scoring. In addition, the analytical evaluation would provide more information on writing traits that may have influenced evaluator judgments. Perhaps analytical scoring would be fine-grained enough to detect improvement in some areas. Moreover, scoring data from an “outside team” would provide a point of comparison. For example, if using analytical scoring, the Singaporean as well as US evaluation teams found no improvement, we would be encouraged to review our teaching practices, comparing our business communication course objectives to our assessment criteria and perhaps instituting some revision. Then again, if the US team awarded substantially lower analytical scores than the Singaporean team, then our expectations for student learning might require adjustment.

As described in more detail below, once this analytical scoring was completed we compared the results, running correlations between the holistic and analytical scores; between the analytical scores themselves; and between the scores awarded by the Singaporean and US teams of evaluators on all the tools. These comparisons raised more questions, leading us to undertake a series of textual analyses of several sub-samples.

Looking at these data from a variety of vantage points, both quantitative and qualitative, provided some preliminary information related not only to our in-house questions, but also to questions raised in the research literature as noted above. In the end, it is clear we were unprepared for where this evolutionary process would lead us, particularly the later discovery regarding the influence of what we came to call “national perspectives” on the evaluation of content.

Below we describe our quantitative and qualitative methods and results separately following the order in which they were performed. We begin with the analytical scoring procedures and results. Then we describe the series of textual analyses, providing examples from student responses to illustrate our conclusions regarding team differences. Finally, we define the national perspectives that account for the different content preferences we observed between the Singaporean and US evaluation teams.

Method for Analytical Scoring

Our core sample, respondents, evaluation teams, and analytical scoring procedures are discussed below followed by the quantitative results.

Core Sample for Analytical Scoring

The core sample was selected from the holistically scored corpus from the initial exploratory study. For the analytical scoring, the responses of most interest to us were those from students whose pre-assessment and post-assessment holistic scores differed markedly either up or down. These would, we believed, provide the clearest picture of the traits our evaluators were rewarding or punishing. Accordingly, we took all those responses from students whose post-assessment holistic scores differed from their pre-assessment by two scoring levels or more (higher or lower). This yielded a core sample of 468 responses, from 234 students.

Respondents

The respondents were upper-level business and accountancy undergraduates taking a business and managerial communication course at the Singaporean university. As such, almost all were Singaporean Chinese (Wong & Phooi-Ching, 2000) although precise figures on ethnic composition are not available due to university policy.

English is an official language in Singapore and is the medium of government and commerce. It is a first or second language for the student respondents. All students in Singapore receive their primary and secondary education through English-medium education and are expected to pass the A-level Examination in English during their final year of high school. Those students whose A-level scores fall in the lower range, as well as foreign students, sit for a test during university enrollment and receive English proficiency assistance in their first year if necessary. English is a fully indigenized language in Singapore, and although the majority of Singaporeans have a mother tongue other than English, government policy since independence has been successful in ensuring that the generation from whom our respondents were drawn are fully bilingual.

Singaporean and US Evaluation Teams

The Singaporean team consisted of seven evaluators, including one of us; the US team included the other one of us and totaled four. As university faculty, all the evaluators had advanced degrees and experience teaching business communication (ranging from 2 to 20 years) including giving feedback on various kinds of student writing and on professional writing in business and engineering firms. Experience with holistic and analytical assessment differed team to team, however. All the US evaluators had over a decade of work with holistic and analytical scoring including with the tools employed for this study; all US team members also had used these tools for

student consultations. By contrast, although some of the Singaporean evaluators were familiar with the analytical tools, none had experience using them for systematic evaluation or consultations prior to this study.

The Singaporean and US evaluation teams were also culturally distinct. The Singaporean team included four locally born Singaporeans; two Australians (including one of the authors), both Singaporean residents; and one Austrian-born long-time resident. The composition of this team reflected the Singaporean university system, which employs full time, long-term foreign talent complementing locals as a means to achieve government educational objectives which are taken very seriously. Foreign talent becomes part of this national agenda. Members of the US team were Caucasian Americans, all US-born. Except for the authors, all the evaluators were paid for their participation in this evaluation.

Analytical Scoring Procedures

For this study the core sample of 468 responses were scored by Singaporean and US teams of evaluators using four analytical tools: task fulfillment, coherence, reasoning units, and error interference (Rogers & Rymer, 2001). To insure that the analytical tools were uniformly administered by the evaluation teams, the authors and one of the US evaluators who was highly experienced with analytical scoring, developed a protocol (available from the authors) based on literature for analytical scoring and practices that have been developed in conjunction with assessment programs in major business schools. Important protocol specifications included insuring that the pre- and post-assessment responses were scored intermittently and randomly following procedures for blind evaluation with monitoring for inter-rater reliability. With the exception of the authors and two members of the Singaporean

team, the evaluators did not know which scenario was used for the pre- and which for the post-assessment nor was this discussed at any time during the evaluation. For the analytical scoring process, all the evaluators, including the authors, were blind to the holistic scores that had been previously awarded. Furthermore, once the protocol was agreed upon there was no communication between the evaluation teams or the authors until the scoring was completed, which took each team several sessions occurring over several weeks.

Results of the Analytical Scoring

To begin, we looked for the degree of change in each evaluation team's analytical scores between pre- and post-assessment, comparing the teams' mean scores (via t-tests) on the responses to each scenario on each analytical tool. We then ran Pearson correlations between tool scores as well as between teams (inter-team) and within teams (intra-team) to discover the degree and areas of agreement and difference. Correlations between the holistic and analytical scores and between scores on the analytical tools allowed us to check scoring validity and to see if the tools were sufficiently related yet distinct enough to function as a set.

Means Scores on Analytical Tools: Singapore & US

Although there was some scoring disagreement between teams as seen in the summary of mean scores by both teams in Table 1 below, the intra-team reliability at both sites was high, suggesting that scoring was consistent within teams and that inter-team discrepancies were due to stable systematic factors rather than random error. (Inter-rater reliabilities for scoring on the four tools fell in the range of 91-96% for the Singaporean team and 94-97% for the US team.)

Table 1

Mean Pre/Post Scores on Analytical Tools

Tool	Evaluation Team	Pre	Post	Increase
Task	Singapore	3.01	3.10	0.09
	US	3.43	3.99	0.56**
	Mean Difference	0.42**	0.89**	
Coherence	Singapore	3.55	3.58	0.03
	US	3.48	3.88	0.4**
	Mean Difference	0.07	0.30**	
Reasoning Units	Singapore	3.38	3.37	- 0.01
	US	3.46	3.84	0.38**
	Mean Difference	0.08	0.47**	
Error Interference	Singapore	3.81	3.86	0.05
	US	4.23	4.32	0.09
	Mean Difference	0.42**	0.46**	

** $p < 0.01$ $n=234$ ($n=232$ US post-assessment)

In terms of inter-team results, comparisons of the Singaporean evaluators' pre- and post-assessment scores on the analytical tools revealed no statistically significant movement on any of the four analytical tools--task, coherence, reasoning units, and error interference. The largest difference awarded by the Singaporean team between pre- and post-assessments was on task, which is intended to measure a writer's ability to correctly identify and appropriately address the purpose of the writing. As can be seen in Table 1, on the six-point scale the mean increase on the task tool between assessments was only 0.09, so we cannot claim that progress in this aspect had been demonstrated by the ratings given by Singaporean team. Increases on the other three tools were even smaller, with the mean score on the reasoning units lower in the post-

than in the pre-assessment. Confirming the holistic scores as described below, analytical scoring by Singaporeans showed no significant change between the pre- and post-assessment.

By contrast, the same responses, scored by the US team showed statistically significant progress on task, coherence, and reasoning units, although the difference in error interference scores between assessments was not significant. Furthermore, in Table 1 it can be seen that the scores given by the US team were significantly higher than those of the Singaporean team on two tools in the pre-assessment (task and error interference) and all four tools in the post-assessment.

Overall, task showed the most significant inter-team disagreement, followed by reasoning units. As shown in Table 2 that follows, this disagreement was paralleled by a drop in the inter-team correlation for these tools. Coherence remained the tool with the greatest agreement between the teams; task the greatest disagreement.

Correlations of Holistic and Analytic Scores

To validate the scoring and test the degree to which the analytical tools were operating as a complementary set, correlations between analytical scores and holistic scores were calculated, as well as intra- and inter-team correlations on the four analytical tools.

All correlations between the analytical scores and holistic scores were significant at $p < .01$, which indicates that the analytical tools were related and elaborated the holistic scoring as anticipated (Rogers & Rymer, 2001). For both teams, task followed by reasoning units correlated most closely with the original holistic scores. Correlations between task and holistic scores ranged from .44 to .52, and between reasoning units and holistic scores 0.31 to 0.49. This was true of scores for both the pre- and post-assessments by both teams of evaluators. The weakest

correlation was between holistic scores and error interference (below 0.3 in all cases), replicating what Rogers and Rymer (2001) found--error appears to be less of a factor in holistic evaluation than task, reasoning, and coherence.

Correlations between Analytical Tools

Correlations between the analytical tool scores are suggestive of the way task, coherence, reasoning units, and error interference function as a set. If the tools are complementary there should be tool-to-tool correlations but not extreme overlap. This is because the tools identify different components of writing rather than overall ability, although some common ground is to be expected, as the traits should also be related and evaluators may not always be able to perfectly differentiate between traits. One trait may affect another--errors disrupt coherence, poor coherence makes reasoning harder to follow, and so on. (Rogers and Rymer 2001).

For both teams, the highest between-tool correlations were for task and reasoning units (.785 and .706 for Singaporean team pre- and post-assessment respectively; .758 and .734 for the US team) followed by coherence-reasoning units and task-coherence. This suggests that in applying the tools both teams saw a close relationship between task fulfillment and reasoning, which is not particularly surprising given the nature of the tasks. Both scenarios required the writer to critique an argument, which makes reasoning strategy integral to task fulfillment. The tight correlation between task and reasoning would become more important later in this study when we discovered that the type of reasoning or framing writers used to report the task influenced the Singaporean and US team evaluators differently.

Between-team differences are shown in Table 2. Here the highest correlation is between reasoning unit scores as measured by the two teams in the pre-assessment. Disagreement was evident on the task tool with Singaporean task scores showing a

higher correlation with the US reasoning units and coherence scores than with the US task scores.

Furthermore, the correlation between task scores was considerably lower in the post-assessment than in the pre-assessment (.167, down from .392). This jibes with the greater number of differing inter-team task scores already noted. Reasoning unit scores in the post-assessment also showed a large drop in inter-team correlation compared with the pre-assessment (.258, down from .555). As task and reasoning scores correlate highly, this suggested that the explanation for the Singaporean and US inter-team differences on task scores might be found in the reasoning of the responses.

Table 2

Pearson Correlations: Singapore & US Inter-team Scores

		Singapore			
		Task	Coherence	Reasoning	Error
Task	Pre	.392**	.264**	.368**	.126
	Post	.167*	.084	.100	.054
Coherence	Pre	.449**	.443**	.404**	.228**
	Post	.322**	.334**	.345**	.286**
Reasoning	Pre	.552**	.481**	.555**	.265**
	Post	.236**	.231**	.258**	.174**
Error	Pre	.147*	.230**	.209**	.251**
	Post	.147*	.138*	.191**	.285**

* p < 0.05 ** p < 0.01

In summary, the student respondents in our sample were consistently rated higher on the analytical tools by US than by Singaporean evaluators. US evaluators' scores show writer improvement following the business communication course,

Singaporean scores did not much as with the holistic scoring. The high correlation between the holistic and analytical scores, which suggests the validity of the scoring overall, also reaffirms that analytical scoring provides information that holistic scoring cannot. At the same time, the differences between evaluation teams was greatest in the area of task, a gap that widened in the post-assessment. The emerging story seemed to be that students had acquired skills that were valued by the US evaluators, but not, apparently, by their Singaporean compatriots. Seeking to understand this finding and its implications for cross-border evaluation, we next looked to the responses or texts themselves.

Textual Analyses of Sub-samples

In order to learn more about the differences between the Singaporean and US teams of evaluators we conducted follow-up textual analyses of actual responses with one analysis leading to the next in a step-by-step process of discovery. In the end, this process involved the selection and analysis of three sub-samples drawn from our core sample: a diversity sample, a task sample, and an outlier sample.

The *diversity sample* consisted of responses from the core sample that had received widely different scores from the Singaporean and US evaluation teams. Totalling 59 responses (16 from the pre- and 43 from the post-assessment), this sub-sample was limited to responses where team-to-team analytical scores differed by two scoring levels or more on any of the four tools.

We, the authors, independently evaluated this diversity sample with the following question in mind: What features in these responses might explain the differences in the analytical scores awarded by the Singaporean and US evaluation teams? We began with quick, independent readings without knowledge of either the holistic or the analytical scores, each taking notes during the reading process.

Upon comparing notes, we agreed that these responses had a lot of similarities (e.g., a "bottom line" structure or an introduction providing the aim and layout of the content to follow). This left us puzzled to explain why the Singaporean and US teams had evaluated these 59 responses so differently. Perhaps inter-team scoring differences could not be attributed to any particular linguistic or rhetorical team preferences?

We followed up with a second independent reading of this diversity sample, this time with knowledge of all the scores. Although this reading did not reveal any clear-cut explanation for the inter-team difference, we agreed not to drop our original line of inquiry too quickly. It was not discrete features that would explain inter-team differences, we concluded, but the differences seemed very likely to have something to do with various content features related to completing the task.

To test this hypothesis we next sought sample responses that were highly significant for their analytical scores on task fulfillment. Not only had content related to task been flagged via our analysis of the diversity sample but also from the analytical scoring results. Recall that on all the analytical tools (task fulfillment, reasoning units, coherence, and error interference) task was the most significant area of inter-team difference--145 of the 234 pre-assessment responses and 182 of the post-assessment responses were given higher scores on task by the US team, with the US mean scores 0.4 higher than those of the Singaporean team for the pre-assessment, increasing to a difference of 0.9 in the post-assessment. According to the US evaluators, the post-responses showed improvement in writing to fulfill the task. Meanwhile, there were only a few cases in the diversity sample in which the scores from the Singaporean team were higher than those of the US team on task, and those mostly in the pre-assessment. Were there common aspects related to reporting the task in these responses that might

explain the high Singaporean scores we wondered? If so, might this also explain the lower US team scores?

It was to responses favored by the Singaporean evaluators on task that we looked next for clues to explain the inter-team scoring differences. These responses might demonstrate what these evaluators valued in terms of content. Since the greatest disagreements between our evaluation teams occurred on task for the post-assessment, it was of the greatest interest here as it would provide a stronger basis for comparison, something we did not have in the diversity sample.

The *task sample* consisted of all those responses in the post-assessment that Singaporean evaluators scored highest on the task tool, responses with scores of 5.0 and above (on a six-point scale with 6 being the highest possible score). We found seven such responses in the 234 post-assessment responses. Independent analysis by the authors revealed that there was a similarity in the type of reasoning employed in these responses. This was not entirely surprising since reasoning is known to correlate highly with task (Rogers & Rymer, 2001). Indeed, these responses reported the information their superior had requested from an assumption of weakness--i.e. The company was at risk of "losing out" unless perceived weaknesses were remedied or concealed from competitors, an observation that is demonstrated later.

To investigate possible team preferences toward reasoning strategies used in reporting the task, we pulled a third sub-sample, an *outlier sample* consisting of responses where the US task scores exceeded the Singaporean scores (or vice versa) by 2 or more scoring levels--eight responses in pre-assessment and six responses in the post-assessment. Five of these 14 responses were favored by the Singaporean team; the US team favored the remaining nine. Reading with knowledge of the scores, we

continued our process of independent analysis with note-taking and subsequent discussion.

Analyses of this third sample confirmed that the reasoning or the framing of the task comprised a principal distinguishing feature between responses favored by US evaluators and those that garnered approval from the Singaporean evaluators. US-favored responses were noticeably less defensive in focus and spoke from a standpoint of maintaining strength; Singapore-favored responses focused on avoiding weaknesses.

Finally, another review of the core sample both tempered and reconfirmed the finding that evaluator team preferences had something to do with the type of framing used when reporting the task. The majority of responses used a mixture of internal and external framing strategies. A preponderance of one type of framing over the other, however, was paralleled by a difference between the two teams analytical scores. Elaboration of these findings is demonstrated with examples below.

Findings from Textual Analyses

Textual analyses revealed the varying framing strategies related to accomplishing the writing tasks that distinguished US-favored from Singaporean-favored responses. An internal (or reactive) framing implied: “Our competitors are a threat, we must remedy our defects to remain viable.” This perspective urged defensive reaction to perceived weaknesses, responsibility for contracts, and a need to protect the organization’s reputation. It appears motivated primarily by fear of failure and need for validation from other parties. Alternatively, external (or proactive) framing implied: “Our company is a threat to our competitors. We have superior expertise to offer, but we need to do a better job showing it.” This approach urged aggressive action to maximize the value of existing assets. The motivation underlying this form of framing appears to be “fear of being overlooked.”

Our textual analyses, particularly of the task and outlier samples, strongly suggested that Singaporean evaluators favored internal framing, whereas US evaluators showed a preference for external framing. As noted earlier, in the core sample as a whole one finds a mixture of these two strategies or little framing at all, as in responses where the task of critiquing is simply not done. Where there was a bias toward one framing approach or the other, however, there was a corresponding imbalance in the task scores awarded by the Singaporean and US evaluation teams. Responses with external, proactive framing got high scores from the US evaluators and low scores from the Singaporeans: when internal reactive framing was employed, the scoring pattern was reversed. Below, we provide some examples from the responses to illustrate this contrast.

Examples from the Pre-assessment

For the pre-assessment the respondent is writing as an assistant to his boss who is the founder of a ball bearing manufacturing company. The respondent is asked to critique and strengthen his employer's argument, which is a sales pitch to another company using customer satisfaction as the selling point. Specifically, the boss has a PowerPoint slide stating that his company has a complaint rate of only 1 in 100 customers. This, he argues, equates to 99% satisfaction with his company's ball bearings. An adequate critique of this slide would involve pointing out the fact that 1 in 100 customers complained does not mean that 99 were satisfied.

Internal Framing for the Pre-assessment

Notice in the Singaporean favored response below how the writer suggests a need to manage the customer's perceptions by rephrasing the number of complaints as a percentage rather than "1 out of every 100" in order to make failure "seem" less likely.

- 1.1 I suggest that we use a percentage to represent 1 out of every 100 customers, as it may seem the chance of having ball bearing failures is more remote.

This response assumes that a 1% chance of failure is high and that failure of the product is therefore likely (although the scenario included no information on industry standards). The use of “may seem” implies the necessity to disguise the possible weakness of the product by clever wording.

In example 1.2, internal focus necessitates that the writer urge his manager to compensate for perceived vulnerability to criticism. The emphasis in the suggested amendment to the boss’s argument is to minimize the impact of the 1% failure rather than touting 99% satisfaction.

- 1.2 I feel that more evidence has to be produced to convince ABC motors that 99% of our customers are satisfied. We could include information from feedback forms completed by customers and evidence of follow-ups by customer service personnel....such fact would be more convincingly prove that IBI customer are indeed satisfied with our ball bearings.

The implication in example 1.2 is clearly that the customer (ABC Motors) might find the figure of 99% satisfaction difficult to believe and will require further evidence. From the necessity to “convincingly prove” customer satisfaction it is likewise assumed that such satisfaction is in doubt. The writer further proposes to show evidence of “follow ups” by customer service personnel, as though the customer might be unwilling to accept the CEO’s word for this. While the 99% satisfaction is mentioned, it appears to be a fragile claim and is presented in terms of potential weakness in the firm’s credibility, rather than as a selling point.

External Framing for the Pre-assessment

In US-favored responses it can be seen that the writer perceives the firm to be in a position of strength with respect to the customer, as in example 1.3 below.

- 1.3 IBI has been in the business of producing motor parts for many years. Over the years, we had built up our reputation of producing reliable motor parts. Our customers now include major automobile manufacturers such as Nissan, Honda, Ford and many others. You could probably used this point to impress onto them the reliability of our products and our reputation.

Here the reputation is established (“...for many years.... we had built up our reputation”), credibility assumed (major clients named), and product quality assured (“impress onto them the reliability of our products”). This is in marked contrast to the internally focused examples above where the reputation is seen to be at risk, the customer’s approval is eagerly sought, and the product’s possible unreliability needed to be concealed.

As may be seen in example 1.4, externally focused responses project an image of confidence that internally focused responses do not.

- 1.4 We should also add depth to the presentation by focusing on the ‘nitty gritty’ and the details. As IBI provides quality products for a diversity of industries, the presentation could follow up with examples of [major clients]. Of course it would be advantageous if we could stress that ball bearings produced by IBI are suitable and reliable for the European as well as the Asian climate and IBI’s standards would not be compromised.

The information presented here reflects a position of strength. It is assumed that greater detail and increased transparency will add to the persuasive appeal--customers should be given details of production and other customers. Strengths are emphasized--the

writer notes that the bearings are suited for European as well as Asian climates. There is no suggestion that the firm's reputation may be threatened by greater frankness. This approach differs markedly from that shown in 1.1, which the Singaporean evaluators preferred and in which the writer is at pains to mitigate the threat of a negative reaction from the customer.

Text Examples from the Post-assessment

In the post-assessment scenario, respondents were asked to critique in the form of counter-argument. The boss had it in mind to scrap previously generous educational benefits which allowed extended study in top universities, in favor of shorter better-targeted courses at less prestigious institutions. This was due to the increasing costs and the point that such generous benefits did not appear to guarantee loyalty. The boss asks the respondent, a junior assistant who might be affected by the proposed change, to draft a counter-argument to his proposal with the idea that such would furnish a kind of critique that he might use to strengthen his stance.

Internal Framing from the Post-assessment

In post-assessment example 2.1, the writer attempts to assist his manager by critiquing the manager's argument in favor of scrapping educational benefits that had previously allowed employees to study at top business schools.

- 2.1 It is of a common concept that traditional graduate studies, especially in top business schools around the world are more prestigious and more comprehensive. The degrees from shorter alternative graduate courses are likely to pale in comparison to those from traditional graduate studies... [You must] convince Ms. Wong and the other senior consultants that these alternative programs not only save costs and time, they are beneficial to the students as traditional programs

The critique in 2.1 should help the manager refine his argument so as to convince “Ms. Wong” and the other directors. The appeal rests entirely on the quality of the “traditional programs,” which the writer uses to make the point that employees denied educational benefits will be disappointed and leave the firm. Unlike the externally focused examples cited below, there is no attempt to recognize the quality of the employees. Rather, the focus draws on organizational deficiencies--e.g., the firm needs employees trained in top business schools to give the organization credibility. The *argumentum ad populum* in the first clause “It is a common concept” suggests the writer’s concept of the company within its competitive environment: it must be cautious of business norms in order to retain credibility. The firm is perceived as shaped by its environment rather than the reverse.

In example 2.2, the writer attempts to counter the argument that less prestigious schools offer better-targeted, non-degree courses that might be a more cost-effective method of training employees.

2.2 You have also suggested that alternative programs offered by less-prestigious schools are generally cheaper and smarter. However, top business schools are well established and have a reputation for producing good students. Hence, students from the top schools may benefit more in their education.

Furthermore, the certificates issued by the top business schools are well recognized, as the public know that these schools provide a high standard and quality of education.

As in example 2.1, here again the argument for cost-effectiveness is answered with an appeal to social credibility. The inference is that the firm will lose the approval of “the public” if certificates from top business schools are not on display.

Another observation concerns the sentence “students from top schools may benefit more from their education.” Here the appeal appears to suggest “social responsibility” rather than strictly business criteria. Such approaches were evident in many responses favored by Singaporean evaluators: the notion that the company had an obligation to deliver on a previously implied promise of education to entrench against critique or forces that might jeopardize the company in some way. This further reinforces the suggestion of weakness as mere individuals within the company. The employees are construed as unable to gain these benefits by other means than by appealing to the boss’s pity.

External Framing from the Post-assessment

In externally focused responses preferred by US evaluators writers did not make lengthy reference to the quality of the major business schools or the possible loss of reputation that might be incurred with a switch to shorter courses. Instead, the persuasive appeal was based on attracting the best recruits as employees. This is to say that the best recruits deserve the best education in order to move up in the firm or they might not want to join, as seen in example 2.3 below.

- 2.3 A simple survey conducted recently by the firm showed that the most attractive benefit that consultants, especially the best applicants look for in our firm is our educational benefit....this educational benefit serves as a good recruitment point for us to attract the best graduates from each cohort. It is an even stronger motivator for those best people to reach the highest capacity....The suggestion to provide alternative programs that are less costly demeans their abilities.

Here the “self-positioning” is slightly different from that in the pre-assessment responses favored by our US evaluators. Here the writer identifies with the best

graduates as the best new hires for the firm, whereas in the pre-assessment the writers identified with the firm with respect to other firms. The position of strength therefore resides with the quality of the new hires with which the writer associates.

In example 2.4 consultants are seen as holding a degree of power over the firm in relationship to attracting and keeping customers “if they were to stop upgrading themselves...” It is further assumed that (a) the consultants presently maintain their education and (b) the company is at present competitive but has the potential to become less so with the implementation of the decision.

2.4 In order to cater to the needs of our customers, consultants need to upgrade themselves. If they were to stop upgrading because of the change in the benefit scheme, Transair will be in a disadvantageous position. We might not be able to provide better and more appropriate services to our customers.

Transair will be less competitive in the industry as a result.

In example 2.4 the firm is portrayed as the loser, whereas in the Singaporean-favored responses like 2.1 and 2.2, emphasis was on the potential personal loss to the employees.

While superficially similar, external and internal foci reflect contrasting views of the writer's in-group, the group on whose behalf the message is composed. In our study the in-group comprises the company in the pre-assessment and the employees in the post-assessment. Internally framed reasoning is reactive, defensive, and under-confident. This is a modest view of company and employee value with an acknowledged need for external validation. By contrast an external focus reflects writer confidence in the in-group. While cooperation of another party might be needed, the in-group does not require validation or confirmation of its value from an external source.

Discussion

Why did Singaporean evaluators' scores show no improvement where the US did, particularly on task? Even with rigorous controls to insure a fair assessment of the writing, might not Singaporean evaluators regard their own compatriots more generously than outsiders? After all, these local evaluators were more accustomed to dialect features in the Singaporean students' written responses to the pre- and post-assessments. Moreover, as instructors, might there be an obvious threat to professional pride from low performing students? The expectation of local bias was not borne out, however. In fact the reverse was the case: on the whole, the US evaluators awarded the highest scores. This was despite the fact that the US university where our evaluation team was based is known as a prestige school where high standards of writing are considered to be a basic requirement of students. What then might account for the difference?

As our textual analyses suggested, differences between the Singaporean and US evaluations can be explained by a preference for the way the content was framed: Singaporean evaluators favoring responses with an internal focus, US evaluators an external one. But what in the cultural or environmental contexts of these evaluation teams might account for these preferences?

A review of historical cultural sources suggested that these evaluator preferences stemmed from what we came to characterize as "national perspectives." By national perspectives we do not mean the much broader concept of culture, but rather an aspect of it. While there is little consensus as to what constitutes an aspect of culture (e.g. compare Usunier, 1996, to Salacuse, 1999), definitions do include the national orientation of groups. For example, Sheer and Chen (2003) explain that "culture is commonly defined as a set of shared and enduring meanings, values, and

beliefs that characterize national, ethnic, or other groups and orient their behavior” (see also Faure & Sjostedt, 1993). National perspective, as aspect of culture, seemed to best characterize what we were observing here.

Below we describe the national perspectives we observed to be operating in our Singaporean and US evaluations, recognizing that these are not static but rather dynamic perspectives being negotiated and renegotiated, particularly in this time of global change and positioning.

National Perspectives Seen in the Responses

The Republic of Singapore is home approaching four million people with Chinese comprising 77.2% of its population, the remainder being Malays, Indians, and a small number of various others (Wong & Phooi-Ching, 2000). Since independence from Great Britain and expulsion from Malaysia in 1965, which is said to have contributed the “ideology of survivalism” that has come to “infuse virtually all aspects of government” (Lawson, 2001, pp. 72 & 73), Singapore has moved from being an economic backwater to a showpiece of Asian capitalism. This prosperity is a great source of pride to Singaporeans but is seen by them as fragile, challenged by their own multi-culturalism, and dependent on constant improvement to keep pace with the rest of the developed world. Since Singapore’s development was so rapid, many Singaporeans alive today know the reality of poverty and the need for struggle to avoid it. As Lawson observed,

[M]otifs linking Singapore’s inherent vulnerability to the threat of ethnic disintegration and the state’s very capacity to survive appear regularly in ministerial and other political speeches. These are often linked with other aspects of government policy, such as the promotion of an ethnically

neutral, pragmatic meritocracy and the building of an all-embracing national identity (2001, p. 64).

The proximity of less fortunate neighboring countries reinforces this knowledge, a state of being that regularly appears in front page articles in Singapore's *Straits Times*.

As Singapore's first Prime Minister, now Senior Minister and elder statesman Lee Kuan Yew remarked recently in a speech to university students:

Much depends on our younger generation, your generation. Do you have the guts and gumption that your parents and grandparents displayed when they faced the stark choice of either working together to make Singapore succeed, or face the humiliation of failure, wishing we had never been ousted from Malaysia? Amid the comfort and affluence of present-day Singapore, do you feel that urge to stay ahead of the pack in order to maintain our lead? You have the advantage of building on the efforts of your elders: educational and economic opportunities among the best in the world, in terms of physical and social infrastructure, travel, education and work abroad to expand your horizons. They give you a precious advantage over others in the region. But the others are hungrier and more driven to get ahead. (Lee, 2003)

Here we see the ideology that drives Singapore. It is a nation that by virtue of its size is not a superpower and in the global economy is competing against larger nations.

Singapore cannot hope to define the global economy, but must work to remain in control of its own destiny.

The US, by comparison, is considerably more secure in its position as the world's remaining superpower. US national ideology may at times seem quite vague given its great diversity, space, and flexibility (Hart, Jennings, & Dixon, 2003).

However, situations of national crisis, such as the attack that destroyed the World Trade

Center, reawaken American nationalism, bringing to the surface the propositions that tie its diverse peoples and regions together. While some may argue that this most recent manifestation prompted a darker side in response, the US perspective is an external one, that of bringing “light to the world.” In recent discourse, the world is presented as a follower rather than a threat that cannot be defended against or a pack that must be outrun. An example of the national rhetoric is seen in a recent presidential address:

Americans are a resolute people who have risen to every test of our time. Adversity has revealed the character of our country, to the world and to ourselves. America is a strong nation, and honorable in the use of our strength. We exercise power without conquest, and we sacrifice for the liberty of strangers.

Americans are a free people, who know that freedom is the right of every person and the future of every nation. The liberty we prize is not America's gift to the world, it is God's gift to humanity. (Bush, 2003)

Our purpose here is not to comment on the merit of these worldviews, but to argue that this national mythos or rhetoric may shape an underlying bias toward a certain type of framing, even when evaluating student writing. In the case of Singaporean evaluators we found a preference for internal focus in reasoning: perceived flaws should be minimized, lest they pose a threat. Reputation must be guarded; external validation and recognition are sought. The in-group is construed as the underdog in its relations with outside parties. Expected standards are set by external others.

US evaluators, by contrast, demonstrated a preference for framing from a perspective of confidence, the task a need to communicate deserved reputation. Flaws in the in-group are seen as less significant than those of comparison groups and can be discussed frankly. Failure is regarded as a consequence of in-group inaction rather than of criteria validated externally. Standards are set by the in-group and are not challenged by comparison with outside parties.

This study suggests that these ideological frameworks influenced evaluators' judgments of writing quality: US evaluators showing a clear preference for those responses that attempted to report using an external focus, Singaporean evaluators either penalizing or failing to reward such an approach and awarding high scores rather to defensive posturing.

Although our focus has been on evaluator judgments, it is interesting to consider the student respondents. Might the higher scores awarded by the US team and the lower scores from the Singapore evaluators suggest that the Singaporean students' framing strategies have become more "American," perhaps as a result of their studying business communication? We cannot make this claim based on evidence. First, it should be remembered that in the core sample as a whole most students used a mixture of external and internal foci with a lesser number showing a clear preference for one over the other. Also, although the US task scores showed an increase between pre- and post-assessment that could be used to argue for learned behavior, the scoring might also be explained by a difference in the tasks themselves, the second task requiring the respondent to argue against the boss's proposed change in company policy rather than to critique the boss's argument as was required for the first.

Still, if there is transference of “Americanized” rhetorical strategies to Singaporean students studying business communication, one might speculate that teaching materials may have played a role. The business communication course at the Singaporean site at the time of this study made use of Locker’s (2000) *Business and Administrative Communication* as a prescribed text, with Guffey’s (2000) *Business Communication, Process and Product* recommended for additional reference. As the Singaporean market is small there is little local material and in all the business disciplines US practices and models tend to be held up as the standard. Thus, when students are taught American norms but assessed by local evaluators there may well be a certain dissonance created between the styles that are taught and those that are rewarded. A generational dimension may also be present: senior university staff in Singapore, including those who participated in this study, went through an education system that was substantially more British than the current one, a system that also influences expatriates of this generation who have joined the system as educators. By contrast, Singaporean students emerged at a time when the age of technology and democracy has turned attention away from Europe to the US.

Future investigations into the influences of national perspectives between Singaporean and US evaluators might elaborate or refute its significance. This follow-up research might also observe the current shift toward Asian values, including democracy based on Confucianism not Locke (Lawson, 2001). Could it be, perhaps, that workplace socialization or maturity in Singaporean society will engender a shift away from framing strategies that may have been seen as arrogant or pushy in earlier days? If there is a growing preference for external focus, is it a reflection of a

changing Singaporean mindset or simply a reflection of individual inexperience that will change when joining the workforce?

As for other future research, interesting questions remain unanswered. First, it is still unclear how the differing judgments of student writing were arrived at. Did evaluators reward one type of framing or punish the other? Second, although we have identified the phenomenon of national perspective, we have no tool for assessing or quantifying it. While various analytical tools (like the Rogers and Rymer reasoning units tool) may be used to score the success of reasoning strategy for a particular readership, an actual repertoire of possible strategies is not provided for the evaluator. Some sort of mechanism to study the framing via quantitative and qualitative analyses, say of Singaporean and US writing seems a logical next step. Future research might also seek to replicate this finding in different contexts. Do business managers share the kinds of evaluator preferences in evidence here? Might workplace communications (e.g., emails or phone conversations with customers say by call center employees) be evaluated from similar perspectives across national boundaries? The perspectives discovered here may provide a starting point for such investigations.

Conclusion

This research suggests that analytical tools targeting a specific population and academic training (e.g. business communication in business schools) can be quite consistently applied *within* evaluator groups but somewhat less so *across* evaluator groups, even when these groups share a disciplinary focus and work in comparable schools in English-speaking countries. Although the US and Singaporean evaluation teams followed the same protocols and scored the same responses, there were

perceptible differences in the mean analytical scores awarded by the two teams, particularly in the post-assessment evaluation of task. The explanation for this may stem from different interpretations of the tool descriptors, different interpretations of the task requirements, or the greater experience with these tools among the US evaluators resulting in an increased willingness to award high scores. Against this view, however, it must be observed that both teams reported high levels of inter-rater reliability, suggesting that any error in the scoring was systematic rather than random. Inexperience or inability to apply the tools correctly would be likely to produce poor reliability and a wide range of interpretations of the tool descriptors, yet this did not occur. Our correlative evidence suggested that interpretation was consistent within teams, if slightly less so between teams. Moreover, textual examination ruled out directness, dialect features, and error as significant factors, suggesting instead that the key determinant in differing evaluator judgments of task and reasoning across teams had something to do with content development.

Generally, this study tends to bear out previous findings that content has the greatest impact on evaluators' scores (e.g., Cumming, Kantor & Powers, 2001; Freedman, 1977, 1979 a & b; Rogers & Rymer, 2000). Error, non-native constructions, and dialect features of Singapore English were not a major source of difference in judgments between our evaluator teams. Instead, differences occurred in judgments of content development related to reporting the task, judgments of quality that seem to coincide with the national perspectives of the evaluators. It would be a mistake, we affirm, to assume that good writing is a stable entity between national contexts despite shared language and democratic ideals and even though evaluators are in the same discipline using evaluation tools targeted for training environments with shared goals. Even with all these contextual factors in common, ideologies related to national

identities may influence evaluators toward quite different determinations, especially when evaluating content development.

This result has implications for international assessments, teaching, and research. For some time research has told us that we cannot assume that evaluators in different countries have the same expectations for the way content should be developed or have the same interpretation of what it means for writers to fulfill an assigned task, even when these evaluators are disciplined via scoring criteria and strict protocols, and even, as we have shown here, when these evaluators share a disciplinary focus and pedagogical goals. One possible explanation may be the fact that evaluating content requires subjective judgments of acceptability for which evaluators may have different thresholds. Complicating the picture, this study suggests that evaluators' judgments of quality may be colored by national perspectives.

The notion that national perspective may impact evaluator judgments across borders may elaborate the suggestion made by Purves (1992) when he declared that the IEA study did not proceed as expected and that "perception" plays a significant role in evaluation. Nationally diverse evaluator teams who shared a common disciplinary focus and related analytical tools produced comparable scoring overall, yet demonstrated different perspectives in rewarding the way the content was framed.

Among the possible implications of this line of research are two. First, the dominance of English as a world business language may create a superficial similarity masking underlying national differences that influence judgments of writing quality. Can we assume that because two people have a language, discipline, and goals in common that they will be influenced by similar arguments or swayed by similar rhetorical strategies? Building on previous research, this study suggests not. While textbooks are replete with advice on intercultural situations, much of it centers on more

obvious differences such as dress, greetings, attitudes toward time, and so forth. Text-level differences are much harder to decipher and therefore to describe as we know from attempts to associate particular preferred strategies in English with cultural orientations (e.g. Sriussadaporn-Charoenngam & Jablin, 1999; Hwang, 1990).

A second implication exists for educators. Like business, education is increasingly globalized. University tests of language skill are evolving to emphasize performance assessment of communicative competency; meanwhile, educators are keen to expose students to the challenges of communicating across borders. We suggest that it is desirable that teachers and evaluators seek heuristics to explore how national perspectives may color judgments. Perhaps the textual examples provided here could even generate some classroom discussions about the impact of national perspectives on framing content.

References

- Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly*, 31, 71-94.
- Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.
- Bizzell, P. (1992). *Academic discourse and critical consciousness*. Pittsburgh: University of Pittsburgh Press.
- Breland, H. M., & Jones, R. J. (1984). Perceptions of writing skills. *Written Communication*, 1, 117.
- Brown, R. L., & Herndl, C. G. (1986). An ethnographic study of corporate writing: Job status as reflected in written text. In Barbara Couture (Ed.), *Functional Approaches to Writing: Research Perspectives* (pp. 11-29). Norwood, NJ: Ablex.
- Bush, G.W. (2003) State of the Union address (January 8 2003). Retrieved August 8, 2003, from <http://www.whitehouse.gov/news/releases/2003/01/20030128-19.html>
- Carson, J. G. (1998). Cultural backgrounds: What should we know about multilingual students? *TESOL Quarterly*, 32, 735-740.
- Chee, C. P. (1996). Rhetoric East and West. In S.A. Abraham, & V. Y. Hsui (Eds.), *The English Language in Singapore Series: Current perspectives on the teaching of writing* (pp.30-43)
- Connor, U. (1987). Argument patterns. In U. Connor & R. B. Kaplan (Eds.), *Writing Across Languages: Analysis of L2 Text* (pp. 57-72). Reading, MA: Addison-Wesley.
- Connor, U. (1990). Linguistic/Rhetorical measures for international student persuasive

- writing. *Research in the Teaching of English*, 24(1), 67-87.
- Connor, U., & Biber, D. (1988). *Comparing textual features in high school student writing: A crosscultural study*. Unpublished manuscript.
- Connor, U., & Lauer, J. M. (1985). Understanding persuasive essay writing: A linguistic/rhetorical approach. *Text*, 5, 309-326.
- Connor, U., & Lauer, J. M. (1988). Cross-cultural variation in persuasive student writing. In A. C. Purves (Ed.), *Contrastive Rhetoric* (pp. 138-159). Newbury Park, CA: Sage.
- Cumming, A., Kantor, R. & Powers, D.E. (2001). Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: an investigation into raters' decision making and development of a preliminary analytic framework. *TOEFL Monograph Series* RM-01-04. Princeton NJ: Educational Testing Service.
- Doheny-Farina, S. (1986). Writing in an emerging organization. *Written Communication*, 3(2), 158-185.
- Faure, G.O. & Sjøstedt, G. (1993). Culture and negotiation: An introduction. In G.O.Faure, & J.Z. Rubin (Eds.), *Culture and negotiation: The resolution of water disputes* (pp. 1-13). London: Sage.
- Freedman, S. (1977). *Influences on the Evaluators of Student Writing*. Unpublished doctoral dissertation. Stanford University.
- Freedman, S. W. (1979a). How characteristics of student essays influence teachers' evaluations. *Journal of Educational Psychology*, 71(3), 328-338.
- Freedman, S. W. (1979b). Why do teachers give the grades they do? [Special Edition]. *College Composition and Communication*, 161-164.
- Ferris, D. R. (1994). Rhetorical strategies in student persuasive writing: Differences between native and non-native English speakers. *Research in the Teaching of*

English, 28(1), 53.

- Gere, A. R. (1980). Written composition: Toward a theory of evaluation. *College Composition and Communication, 33(4)*, 45-58.
- Gorman, T. P., Purves, A. C., & Degenhart, R. E. (1988). *The IEA study of Written Composition, 1*. Oxford: Pergamon Press.
- Guffey, M.E. (2000). *Business communication: process and product (third edition)*. Cincinnati: South-Western College Publishing.
- Harris. W. (1977). Teacher response to student writing: A study of the response patterns of high school English teachers to determine the basis for teacher judgment of student writing. *Research in the Teaching of English, 11*, 175-185.
- Hart, R.P., Jennings, W.P., & Dixson, M.J. (2003). Imagining the American people: Strategies for building political community. *Journal of Communication, 53 (1)*, 138-154.
- Hayes, J. R., & Hatch, J. A. (2000). Does holistic assessment predict writing performance? *Written Communication, 17(1)*, 3-26.
- Haswell, R. H. (1998). Rubrics, prototypes and exemplars: Categorization theory of writing placement. *Assessing Writing, 5*, 231-268.
- Herrington, A. J. (1985). Writing in academic settings: A study of the contexts for writing in two college chemical engineering courses. *Research in the Teaching of English, 19(4)*, 331-361.
- Hwang, J.R. (1990). 'Deference' versus 'politeness' in Korean speech. *International Journal of Sociology of Language, 82*, 41-55.
- Kubota, R. (1999). Japanese culture constructed by discourses: Implications for applied linguistics research and ELT. *TESOL Quarterly, 33*, 9-35.

Lawson, S. (2001) Sanitizing ethnicity: the creation of Singapore's apolitical structure.

Nationalism & Ethnic Politics, 7 (1), 63-81.

Lee, K.Y. (2003) Speech to Nanyang Technological University students (18 February

2003). Retrieved August 8, 2003, from

http://www.getforme.com/previous2003/previous230203_SMLeeKuanYewSpeaksToUniversityStudents.htm

Lloyd-Jones, R. (1987). Tests of writing ability. *Teaching Composition*, 12, 155-176.

Locker, K.O. (2000). *Business and administrative communication (Fifth Edition)*.

Boston: Irwin McGraw-Hill.

Lunsford, K. J. (2002). Contextualizing Toulmin's Model in the Writing Classroom.

Written Communication, 19(1), 109-174.

McCloskey, D. (1985). *The rhetoric of economics*. Madison University of

Wisconsin.

Paradis, J., Dobrin, D., & Miller, R. (1985). Writing at Exxon ITD: Notes on the

writing environment of an R&D organization. In L. Odell and D.L. Goswami

(Eds.), *Writing in nonacademic settings* (pp. 281-307). New York: Guilford.

Purves, A. C. (1992). Reflections on research and assessment in written composition.

College Composition and Communication, 33(4), 108-122.

Purves, A. C., Groman T. P., & Takala, S. (1988). The development of scoring scheme

and scales. In T. Groman, A. C. Purves, & R. E. Degenhart (Eds.), *The IEA's*

study of written composition I: The international writing tasks and scoring

scales, 5 (pp. 41-58). New York: Pergamon.

Purves, A.C. & Takala, S. *An International Perspective on the Evaluation of Written*

Composition Evaluation in Education: An International Review Series [Special

Issue] 5, (2).

- Raforth, B. A., & Rubin, D. L. (1984). The impact of content and mechanics on judgments of writing quality. *Written Communication, 1*, 446-458.
- Ramanathan, V., & Kaplan, R. B. (1996a). Audience and voice in current composition tests: Some implications for ESL student writers. *Journal of Second Language Writing, 5*, 21-34.
- Ramanathan, V., & Kaplan, R. B. (1996b). Some problematic “channels” in the teaching of critical thinking in current L1 composition textbooks: Implications for L2 student writers. *Issues in Applied Linguistics, 7*, 225-249.
- Raymond, J. C. (1982). What we don't know about the evaluation of writing. *College Composition and Communication, 33(4)*, 399-403.
- Rogers, P. S. (1989). Choice-based writing in managerial contexts: The case of the dealer contact report. *Journal of Business Communication, 26(3)*, 197-216.
- Rogers, P. S. (1994). Analytical measures for evaluating managerial writing. *Journal of Business and Technical Communication, 8(4)*, 380-407.
- Rogers, P. S., & Rymer J. (1996). AWA Analytical Tools Set: Analytical Writing Assessment Diagnostic Program. *Report to Graduate Management Admissions Council*. McLean, VA: USA
- Rogers, P. S., & Rymer J. (2001). Analytical tools to facilitate transitions into new writing contexts. *The Journal of Business Communication, 38(2)*, 112-152.
- Rogers, P.S. (2003). *Organizational Voice: Explicit factors enabling assimilation and participation*. University of Michigan Working Paper #03-001.
- Salacuse, J.W. (1999). Intercultural negotiation in international business. *Group Decision and Negotiation, 8*, 217-236.
- Schrivier, K. A. (1989). Evaluating text quality: The continuum from text-focused to reader-focused methods. *IEEE Transactions on Professional Communication,*

32(4), 238-225.

- Sheer, V.C. & Chen, L. (2003). Successful Sino-Western business negotiation: Participants' accounts of national and professional cultures. *The Journal of Business Communication*, 40(1), 50-85.
- Shepherd, G. J., & O'Keefe, B. J. (1984). The relationship between the development level of persuasive strategies and their effectiveness. *Central States Speech Journal*, 35, 137-152.
- Siegal, H. (1997). *Rationality redeemed?* New York: Routledge.
- Sriussadorn-Charoenngam, N. & Jablin, F.M. (1999). An exploratory study of communication competence in Thai organizations. *The Journal of Business Communication*, 36, 382-418.
- Swales, J. M. (1998). *Other floors, other voices: A textography of a small university building*. London: Lawrence Erlbaum Associates.
- Sypher, B. D., Russo, T. C., & Hane, A. C. (2002). The development of persuasive abilities in college. *American Communication Journal*, 5(3). Retrieved August 20, 2003, from <http://acjournal.org/holdings/vol5/iss3/special/sypher.htm>.
- Usunier, J.-C. (1996). Cultural aspects of international business negotiations. In P.N. Ghauri and J.-C. Usunier (Eds.), *International business negotiation* (pp. 91-118). Oxford, U.K: Pergamon.
- Winsor, D. A. (1996). *Writing like an engineer: A rhetorical education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Winsor, D. A. (1999). Genre and activity systems: The role of documentation in maintaining and changing engineering activity systems. *Written Communication*, 16(2), 200-224.

Wong, I.F.H. & Phooi-Ching, L. (2000). Chinese cultural values and performance at job interviews: A Singapore perspective. *Business Communication Quarterly*, 63(1), 9-22.

Zamel, V. (1997). Towards a model of transculturation. *TESOL Quarterly*, 31, 341-352.