

Are We Having the Effect We Want? Implementing Outcomes Assessment in an Academic English Language-Support Unit Using a Multi-component Approach

Li-Shih Huang

ABSTRACT

In response to a call by writing-center researchers and professionals since the 1980s for rigorous scientific assessment, this paper reports on a project that is the first to implement a multi-component assessment of a writing center based on a year-long gathering of data from multiple sources, including the writing center's usage profile, users' perceptions related to writing, learning outcomes-based evaluations, and a satisfaction survey. In addition to the value of the findings from this project and its implications for writing-center research and practice, information about the procedures involved in implementing this multi-component assessment approach may be useful to administrators, researchers, and practitioners in academic language-support units across institutions of higher education.

INTRODUCTION

In response to the call since the 1980s for rigorous scientific assessment issued by writing-center researchers and professionals, this paper reports on a writing-center-assessment project that is the first to implement a multi-component assessment with a year-long (academic year 2009–2010) gathering of data from multiple sources, including the Writing Center's usage profile, users' perceptions related to writing, learning outcomes-based evaluations, and a satisfaction survey. A previous paper detailed the component centered on a center usage profile constructed as a source of outcomes assessment.¹ This paper reports on the results from the three remain-

ing components: a learning outcomes-based assessment, users' perceptions related to writing, and a satisfaction survey, all conducted for the purpose of critically evaluating the effectiveness of academic English language support and services provided by a writing center at a Canadian university. In addition to the value of the findings from this project and its implications for writing center research and practice, the information about the procedures involved in implementing this multi-component assessment approach may be useful to administrators, researchers, and practitioners in academic language-support units across institutions of higher education.

REVIEW OF RELEVANT RESEARCH

Although writing centers sometimes resist assessment for immediate practical and longer-term implicational reasons, it has proven to be beneficial both for evaluating the effectiveness of services provided by academic English language-support units in order to plan and improve services to students and for answering the age-old question: Does what we do matter (Niller; Henson and Stephenson)? In response to the call since the 1980s for rigorous scientific assessment issued by both the research and practice-oriented communities in the field of writing-center research (Hawthorne; Henson and Stephenson; Lerner; Neuleib), this paper reports on a writing-center outcomes assessment project in the context of a newly established language-support unit.

Since the 1970s, studies have used numerous methods to substantiate the impact or effectiveness of writing centers. These methods have included, for example: focus groups (e.g., Cushman et al.), comparison of grade-point averages or course grades (e.g., Lerner; Newmann; Roberts; Sandlon; Sutton and Arnold; Waldo), surveys (e.g., Morrison and Nadeau), proficiency exams (e.g., Naugie; Hyland, Howell, and Zhang), comparisons of drafts before and after students' visits to the writing center (e.g., Bennett; David and Bubloz; Henson and Stephenson; Niller), using an error recognition test (e.g., Wills), students' satisfaction levels (e.g., Carino and Enders), measurement of students' attitudes toward writing (e.g., Ady; Clark; Davis; van Dam), and students' perceived levels of confidence after visiting the writing center (e.g., Ady; Matthews).

Writing-center researchers have emphasized the need for an evidence-based approach to outcomes assessment (e.g., Bell; Hawthorne; Henson and Stephenson; Pemberton) that is critical to moving the field forward (e.g., Lerner). Since the 1980s, many have cited such issues as time and resource constraints, the need for expertise in assessment research methods, (mis-)conceptions about purposes of assessment (e.g., Lerner; Schuh

and Upcraft), and, most importantly, difficulties involved in substantiating the link between the support received and any improvement in students' writing (see Jones; Pemberton and Kinkead) as major challenges. Writing-center researchers have developed a rich body of qualitative work, and, in recent years, the field has also witnessed efforts to utilize quantitative methodologies to assess writing centers' work (e.g., Henson and Stephenson; Niller); however, to date, such evaluation studies are still lacking (Hawthorne; Jones).

Most qualitative studies that provide indirect evidence of students' writing-ability improvement have indicated writing centers' positive impact, but the results have been less clear-cut in quantitative studies. Some have found significant improvements in writing produced by students who use writing center (e.g., David and Bubloz; Henson and Stephenson; Carino and Enders; Sutton and Arnold), but others have indicated inconclusive results (e.g., Bennett; Roberts). There have also been cautions issued about the use of statistics (e.g., Enders; Lerner). As many have pointed out, the field is rich with anecdotal evidence, reflections, and studies that are qualitative in nature, but quantitative studies published in journals remain lacking (e.g., Jones; Johaneck; Lerner; Neuleib). Although I agree with Johaneck, Lerner, Jones, and others who call for quantitative methodologies, I argue that the types of questions asked should drive the methods that one employs, whether they are be quantitative, qualitative, or mixed, to fully address the questions at hand (Tashakkori and Teddlie).

Ultimately, assessment in the context of higher education and in the case of a language-support unit is a process of asking questions about what students have gained from the center (e.g., knowledge or information, thinking, or performance-based abilities). A variety of proposed assessment or program-evaluation models have various orientations in line with the purpose of the assessment (e.g., Hawthorne; Upcraft and Schuh). According to Allen, assessment may involve asking questions about "students' satisfaction with their educational experience," "the amount of their engagement or participation," and/or "what they actually gained from that experience" (96).

The current project, drawing on Allen's work and that of Schuh and Upcraft regarding the student-services assessment model, is an outcomes-oriented assessment that is complemented by non-outcome data to present a fuller picture of the program. The categories of outcomes, as also clarified by Allen, involve the following: For engagement or participation-based assessment, it involves analyzing how the center is being utilized over a period of time. In the questions for a satisfaction-based assessment, the aim is to learn what users did and did not like, or how they felt they did or did

not benefit as learners by asking students to rate their satisfaction with, for example, the quality of tutoring, the variety of offerings, the support anticipated and received, and the perceived benefits or impact on their academic work. Research investigating questions about learning outcomes, which is the component that is often missing or lacking in writing-center research, deals with the crucial issue: What can the student do as a direct result of the support received that he/she did not demonstrate before receiving the support?

This study implemented a multi-component assessment strategy by gathering both direct and indirect evidence from multiple sources over the course of an academic year. The goal was to provide information that would help more language-support units engage in and benefit from systematic self-examinations of their programs. Specifically, this paper addresses the following research questions:

1. *Users' writing-related perceptions:* What are users' perceptions about their writing after their visits to the Writing Center?
2. *Learning outcomes:* Is there a difference between the drafts produced by students before and after their Writing Center visits? Is there a difference, as indicated by students' writing, between Writing Center users and nonusers?
3. *Satisfaction survey:* How satisfied are students with the offerings and support provided by the Writing Center?

THE INSTITUTIONAL CONTEXT

The outcomes (or more accurately, progress) assessment project was undertaken at a mid-size, comprehensive university in Canada. The university's academic programs include ten faculties and two major divisions. According to the enrollment figures in 2008/09, approximately 8% of its 22,025 undergraduates and 13% of its 2,593 graduate students are international students. In 2006, the university founded an English Language Proficiency Working Group to examine policies supporting and challenges facing students. In 2007, the university's Writing Center was established as part of the Learning and Teaching Center, which serves the academic language-learning needs of both graduate and undergraduate students. In 2007, a needs-assessment research project was undertaken to better evaluate graduate and undergraduate students' academic language-learning needs and to review any skill gaps identified by students and instructors (Huang, *Graduate, Seeing Eye to Eye*). The needs assessment was conducted for both English-as-an-additional-language (EAL) and English-as-a-first-language

(EL1) students. Results from the needs assessment were used to inform the development and offering of English for Academic Purposes programs and workshops tailored to both graduate and undergraduate students' needs. This project then led to the current progress assessment conducted in the 2009–2010 academic year, which was designed to evaluate the effectiveness of the support and offerings provided by the Writing Center.

METHODS

Data Collection

The project² employed a multi-component strategy of assessment that involved collecting multiple forms of data. According to standards of high-quality research, triangulation (i.e., using more than one method to approach a topic or situation and ferret out information) is substantially preferable in terms of the resulting quality of information and the trustworthiness that is placed in that information (Denzin). The three major assessment components covered in this paper are as follows:

1. Learning outcomes-based evaluation: The learning outcomes-based assessment examined drafts produced by students who provided their informed consent to participate in the study. An open letter was sent to invite instructors' participation in the study. The letter sought to identify instructors who would be receptive to providing their students the option of visiting the Writing Center before the final grade for each of their papers was assigned. A class visit was subsequently made, with the instructor's permission, to invite students who expressed a willingness to participate in the study and to obtain their consent to collect their written work throughout the term. For every assignment, after submitting their first drafts, students were given one week to voluntarily visit the Writing Center before submitting their second drafts. Clean copies of the first and second drafts were collected through voluntary submissions from both users and nonusers who agreed to participate in the study.

2. Writing- and writing-center-related perceptions: A survey was distributed at the end of the fall term to a group of randomly selected users after their visits to the Writing Center. The survey examined whether the support that users received had helped them in twenty different specific areas (e.g., I have learned strategies for revising; I have learned strategies to continue improving my writing on my own; my ability to discuss issues in class has improved). The purpose was to provide an indication of users' perceptions about their own writing/speaking after receiving support from the Writing

Center, and also of the relation between participants' usage of the writing center and their self-perceptions about their writing/speaking abilities.

3. Satisfaction-based evaluation: A detailed student-satisfaction survey regarding the Writing Center was distributed through the university's registrar at the end of the spring term. The survey, modified from Morrison and Nadeau's instrument, contained five key sections, namely, (a) the respondent's background (e.g., year of study, program of study, department, language background, and gender), (b) questions about the Writing Center's offerings (e.g., frequency of visits, types of support used, and workshops attended), (c) support received and satisfaction levels, (d) tutors and workshop satisfaction, and (e) future involvement with the Writing Center. The survey was designed to gather both qualitative and quantitative data that would address the question related to the extent of students' satisfaction with offerings provided by the Writing Center. Given that the main goal of this assessment project was driven by the needs of a particular program, a balance was sought between generating high-quality information from the survey and maximizing the number of users who would complete a detailed 20- to 30-minute survey with multiple sections.

Data Analysis

In the process of addressing the research questions in each component of the project presented above, both qualitative and quantitative methods that complement each other were utilized to maximize the data obtained. For the learning-outcomes analysis, students self-selected whether or not they would use the Writing Center. All 338 before- and after-drafts, 274 produced by writing-center-users and 64 by nonusers, were collected throughout an academic term. The drafts gathered were each assigned a number code that concealed any identification information such as names, course numbers, submission dates, and so on. To minimize biases that might occur if drafts were placed in sequence as how they were submitted (i.e., the before-draft comes before the after-draft, assignment 1 comes before assignment 2, and so on), all drafts were arranged in random order. The drafts were then rated independently by three raters³ using two scoring methods: analytically for macro-level features (e.g., purpose, development, and organization) and micro-level features (e.g., citation/format, grammar, and punctuation), using scoring rubrics adapted from standardized scoring rubrics suitable for the study (see Appendix A), as well as holistically for the drafts' overall quality.

To address questions regarding differences between the before- and after-drafts and the drafts produced by users and nonusers, the ratings were subjected to statistical analyses using *t*-tests to identify any differences between the before- and after-drafts produced by those who used the writing center and between users of the writing center and those who opted out. Trend-line analyses were conducted for users who submitted three or more drafts over the course of the term to see whether their work improved over time. Finally, both inter-rater and intra-rater reliability were evaluated using the intra-class correlation to ascertain the level of consistency within, between, and among individual raters' analytical and holistic ratings. All analyses were conducted using SPSS Version 15.0, with an alpha level of .05 to determine significance.

RESULTS

1. Learning Outcomes-Based Evaluation

A total of 338 drafts were collected from students across disciplines (e.g., engineering, English, psychology, linguistics, political science, and law). Among those, 64 drafts were produced by Writing Center nonusers, and 274 before- and after-drafts were collected from students who volunteered to use the center.

Before examining the data to answer the research questions related to this component, both inter- and intra-rater reliability were assessed. Results from the inter-rater reliability assessment derived from the ratings of 274 drafts indicated that, for the macro features of purpose, development, and organization, the correlation coefficients were $r = .37$ ($p = .000$), $r = .50$ ($p = .000$), and $r = .42$ ($p = .000$), respectively, and for the micro features of citation/format, grammar, and punctuation, the coefficients were $r = .11$ ($p > .05$), $r = .45$ ($p = .000$), and $r = .29$ ($p = .000$), respectively. The correlation coefficient for overall ratings among the three raters was $.41$ ($p = .000$). The evaluation of development had the highest correlation among the raters; raters' consistency in their ratings of citations/format and grammar was the lowest. None of the correlations exceeded .45, however. When the correlations between raters (i.e., rater 1 vs. rater 2; rater 2 vs. rater 3; rater 1 vs. rater 3) were examined, the values ranged from .29 to .32 for purpose, .30 to .45 for development, .27 to .40 for organization, .25 to .29 for citation/format, .21 to .31 for grammar, .23 to .29 for punctuation, and .31 to .37 for overall ratings at $p = .000$. Overall, the agreement between raters 2 and 3 was slightly higher than that between raters 1 and 3, and raters 1 and 2 had the lowest agreement rate.

Results from the intra-rater reliability assessment derived from the ratings of a subset of 80 randomly selected drafts showed that, for the macro features of purpose, development, and organization, the coefficients were $r = .62$ ($p = .000$), $r = .72$ ($p = .000$), and $r = .69$ ($p = .000$), respectively, and, for the micro-level features of citation/format, grammar, and punctuation, the coefficients were $r = .95$ ($p = .000$), $r = .46$ ($p = .003$), and $r = .37$ ($p = .019$), respectively. On the basis of the results from the inter- and intra-rater reliability assessment, an examination of ratings by each individual rater was conducted in order to gauge the differences in work produced by users before and after their visits to the center.

1.1 A comparison of users' before and after drafts: Is there a difference between the drafts produced by students before and after their Writing Center visits?

As Table 1 shows, overall, ratings for the after-drafts were slightly higher than those for the before-drafts in terms of all macro- and micro-level features, as well as in terms of the scores for the overall rating; this difference indicates that users' writing improved after their visits to the Writing Center. Levene's test for the homogeneity of variances was used to test whether the variances in means of ratings between the before- and after-drafts differed significantly. Results showed that none of the variances were significantly different between the before- and after-drafts, and thus equal variances can be assumed. The t -tests were then conducted to examine the equality of means, and results indicated that none of the differences were significant ($p > .05$).

Table 1. Overall Comparison between Before- and After-Drafts

Level/Feature		Draft	<i>M</i>	<i>SD</i>
Macro	Purpose	Before	3.73	.68
		After	3.83	.60
	Development	Before	3.51	.68
		After	3.78	.62
	Organization	Before	3.63	.69
		After	3.88	.59

Huang / Are We Having the Effect We Want?

Level/Feature		Draft	<i>M</i>	<i>SD</i>
Micro	Citation/Format	Before	3.47	1.00
		After	3.87	.90
	Grammar	Before	4.02	.45
		After	4.17	.37
	Punctuation	Before	4.27	.33
		After	4.33	.30
Overall Holistic Rating		Before	3.53	.66
		After	3.77	.56

Note. $N = 137$ for each before and after drafts (total $N = 274$).

When each rater was examined separately, the results showed that each individual rater rated the after-drafts as better than the before-drafts in both analytical and holistic ratings. To determine whether the mean differences were statistically significant, the means of the analytical features and holistic ratings were compared using *t*-tests. The results showed that, for rater 1, the macro-level features development ($p = .000$) and organization ($p = .000$), the micro-level feature grammar ($p = .019$), and the overall holistic rating ($p = .000$) were significant. For rater 2, the differences between the before and after drafts were statistically significant in terms of one macro feature, organization ($p = 0.43$) and two micro-level features, citation/format ($p = .040$) and grammar ($p = .006$). For rater 3, differences were significant for two macro-level features, development ($p = .029$) and organization ($p = .041$); one micro-level feature, citation/format ($p = .018$); and the overall holistic rating ($p = .008$). The shaded cells in Table 2, which indicate that the differences between the before- and after-drafts were statistically significant, show that organization is the only feature shared among all three raters.

Pearson correlational analyses were conducted to ascertain the relationships between the macro- and micro-level features in the analytic ratings and the holistic ratings. As shown in Table 3, overall, all features in the analytic ratings were correlated with the overall holistic ratings, with very strong correlation values ranging from .821 to .926 between features at the macro level and the overall holistic ratings, and moderate to low correlation values ranging from .410 to .506 between the three micro-level features and the overall holistic ratings. This pattern suggests that raters may have placed more emphasis on macro-level features when determining the holistic scores. Correlations among individual raters also indicated that the

Table 2. A Comparison of Before- and After-Drafts by Individual Raters ($N = 274$)

Rater	Draft	Macro-level Features			Micro-level Features				Overall
		P	D	O	C/F	G	Pu		
1	B	$M = 4.03,$ $SD = .75$	$M = 3.64,$ $SD = .81$	$M = 3.88,$ $SD = .87$	$M = 3.44,$ $SD = .81$	$M = 3.82,$ $SD = .54$	$M = 3.95,$ $SD = .44$	$M = 3.78,$ $SD = .75$	
	A	$M = 4.15,$ $SD = .61$	$M = 4.08,$ $SD = .64$	$M = 4.22,$ $SD = .67$	$M = 3.64,$ $SD = .67$	$M = 3.98,$ $SD = .53$	$M = 3.97,$ $SD = .46$	$M = 4.01,$ $SD = .53$	
2	B	$M = 2.96,$ $SD = 1.23$	$M = 3.01,$ $SD = 1.15$	$M = 2.99,$ $SD = 1.17$	$M = 3.45,$ $SD = 1.6$	$M = 3.83,$ $SD = .68$	$M = 4.22,$ $SD = .49$	$M = 2.92,$ $SD = 1.12$	
	A	$M = 3.12,$ $D = 1.17$	$M = 3.20,$ $SD = 1.11$	$M = 3.25,$ $SD = 1.1$	$M = 3.88,$ $SD = 1.26$	$M = 4.02,$ $SD = .56$	$M = 4.32,$ $SD = .42$	$M = 3.14,$ $SD = 1.06$	
3	B	$M = 4.15,$ $SD = .81$	$M = 3.89,$ $SD = .62$	$M = 4.04,$ $SD = .74$	$M = 4.34,$ $SD = .77$	$M = 4.37,$ $SD = .65$	$M = 4.63,$ $SD = .54$	$M = 3.91,$ $SD = .79$	
	A	$M = 4.28,$ $SD = .71$	$M = 4.09,$ $SD = .71$	$M = 4.21,$ $SD = .66$	$M = 4.63,$ $SD = .641$	$M = 4.51,$ $SD = .54$	$M = 4.69,$ $SD = .478$	$M = 4.14,$ $SD = .65$	

Note. B = before; A = after; P = purpose; D = development; O = organization; C/F = citation/format; G = grammar; Pu = Punctuation. Shaded cells indicate that the differences between the before- and after-drafts were statistically significant at $p < .05$.

Table 3. Correlations between the Holistic Ratings and Features in the Analytic Ratings

Rater	Macro-level Features				Micro-level Features			
	P Mean	D Mean	O Mean	C/F Mean	G Mean	Pu Mean		
1	Holistic Mean	.585**	.774**	.720**	.551**	.497**	.362**	
	<i>p</i>	.000	.000	.000	.000	.000	.000	
2	Holistic Mean	.911**	.948**	.944**	.427**	.401**	.373**	
	<i>p</i>	.000	.000	.000	.000	.000	.000	
3	Holistic Mean	.792**	.867**	.821**	.357**	.192**	.098	
	<i>p</i>	.000	.000	.000	.000	.001	.107	
Overall	Holistic Mean	.821**	.926**	.903**	.410**	.506**	.428**	
	<i>p</i>	.000	.000	.000	.000	.000	.000	

Note. B = before; A = after; P = purpose; D = development; O = organization; C/F = citation/format; G = grammar; Pu = Punctuation. ** Correlation is significant at the 0.01 level (2-tailed).

highest correlations were between development (a macro-level feature) and the holistic ratings for each of the three raters.

1.2 A comparison between users and nonusers: Is there a difference, as indicated by students' writing, between Writing Center users and nonusers?

The results indicated that, overall, the mean ratings of Writing Center nonusers were higher than those of users for all three macro-level features (purpose, $M = 3.20$ vs. $M = 3.00$; development, $M = 3.17$ vs. $M = 3.08$; organization, $M = 3.20$ vs. $M = 3.09$) and for one micro-level feature (citation/format, $M = 4.30$ vs. $M = 3.48$), but lower than those of users for two micro-level features (grammar, $M = 3.82$ vs. $M = 3.95$; punctuation, $M = 4.19$ vs. $M = 4.29$). Overall ratings also indicated that the mean rating of nonusers' drafts was higher than that of the users ($M = 3.30$ vs. $M = 3.13$). With the exception of the micro-level feature citation/format ($p < .001$), the t -test results indicated that there were no statistically significant differences between the drafts produced by users and nonusers of the Writing Center in terms of individual-rated features or their overall ratings.

1.3 Users' development over time: Is there a difference over time between the drafts produced by students before and after their Writing Center visits?

To explore differences between before- and after-drafts over time, repeated measures, which can provide an omnibus test of mean differences over a set of time points, were used to analyze the variables measured over the term. The results indicated a statistically significant difference for the macro feature of purpose ($p = .001$).

A trend-line analysis was conducted to assess the ratings in terms of the presence of trend components. This analysis helped determine which models/lines were the best fit for the ratings, and whether there was any general tendency for the means of the features to increase steadily as time passed. The comparisons were between the mean ratings for the successive assignments submitted over the course of a term. Table 4 presents results from the tests identifying significant trends for the before- and after-drafts, as well as overall trends. Overall, the results revealed that, for the features of organization, citation/format, grammar, and the overall holistic ratings, there were statistically significant linear trends (shaded cells). This finding means that there is statistically significant tendency for the mean level of these features to increase (i.e., move upward) or decrease (i.e., move downward) over time. Results that indicated significant quadratic trends suggested that the outcomes were uneven and that the cubic trend represents a more complex pattern, with more than one change of direction over the course of the

Huang / Are We Having the Effect We Want?

term. None of the features or the overall ratings in the after drafts showed a significant linear upward trend. The trend-line analysis over the course of the term, however, indicated that only one macro-level feature, development, and one micro-level feature, citation/format, had a linear, upward-progression trend ($p < .05$).

Table 4. Results from the Trend-Line Analyses ($N = 274$)

Level	Feature	Draft	Trend	Direction	p
Macro	Purpose	B	Non-linear (Quadratic)	Upward	.002
		A	-	-	<i>n.s.</i>
		Overall	Non-linear (Quadratic)	Upward	.002
	Development	B	Non-linear (Quadratic)	Upward	.000
		A	-	-	<i>n.s.</i>
		Overall	Non-linear (Quadratic)	Upward	.001
	Organization	B	Non-linear (Quadratic)	Upward	.001
		A	Linear	Downward	.011
		Overall	Linear	Downward	.002
Micro	Citation/Format	B	Linear	Upward	.024
		A	-	-	<i>n.s.</i>
		Overall	Linear	Upward	.003
	Grammar	B	Linear	Downward	.000
		A	Linear	Downward	.004
		Overall	Linear	Downward	.000
	Punctuation	B	Non-linear (Cubic)	Downward	.040
		A	Linear	Downward	.047
		Overall	Non-linear (Cubic)	Downward	.017
Overall Holistic Rating	B	Non-linear (Quadratic)	Upward	.000	
	A	-	-	<i>n.s.</i>	
	Overall	Linear	Downward	.004	

2. Writing- and Writing-Center-Related Perceptions

Of the 252 students who provided their email addresses when they visited the Writing Center during the fall term, 72 users responded to the survey focused on their writing- and center-related perceptions after their visits to the Writing Center, for a response rate of 28.6%. Table 5 presents the demographics of users who responded to the survey.

Table 5. Demographic Profile of Participants ($N = 72$)

Level of Study	Undergraduate 37 (51%)		Graduate 35 (45%)	
Divisions	Humanities 14%	Social Sciences 72%	Physical Sciences 7%	Life Sciences 8%
Language	EL1 62%		EAL 38%	
Gender	Male 11%		Female 89%	

Students reported that they visited the center for 39 different courses. When respondents were asked how they learned about the Writing Center, 48.6% reported that they learned by word of mouth, 37.8% were recommended by professors, 21.6% from the website, and 13.5% from emails sent out by the Learning and Teaching Center. Some others learned from beginning-of-the-term-orientations, flyers, or from passing the Writing Center office.

In terms of respondents' frequency of visits, 29.7% visited the Writing Center once during the academic year; 43.3% visited the center between two and five times; 12.8% visited from five to 10 times, and 10.1% visited more than 10 times. The top-five reported reasons for visiting were the following:

1. I want some feedback on an early draft: 73%
2. I need help with grammar in a piece of writing: 67.6%
3. I need help with writing discipline-specific writing: 45.9%
4. I am not sure what to do with a written assignment: 32.4%
5. I need a better grade: 29.7%

According to respondents' comments, users viewed the function of the center as focused mainly on editing and surface issues. In addition, only 8% each of the respondents indicated that they would use the Writing Center for workshops related to oral communication and grant applications. These numbers indicated that students were missing opportunities to view the center as a valuable resource for developing other language domains and specificities, as identified in previous needs-assessment studies (Huang, *Seeing Eye to Eye*).

Table 6 presents the percentages of respondents who responded "agree" or "strongly agree" in the 20 areas related to their perceptions about their writing/speaking. Overall, 78.5% of respondents agreed or strongly agreed that the Writing Center had helped them with their writing-related skills (items 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, and 12). For speaking-related skills, 26.2% of the respondents perceived their speaking skills and confidence had improved (items 4, 13, 14, 15, and 16). A total of 44% of respondents agreed or strongly agreed that workshop(s) helped them perform better in the course; this agreement was corroborated by the finding that 42% agreed or strongly agreed that their course grades had improved. Overall, 58% of respondents agreed or strongly agreed that the Writing Center had helped them in their academic studies. The results in relation to specific individual items showed higher percentages of agreement concerning areas in the writing domain perceived by the students, such as skills learned, enhancement in confidence levels, improvement in ability, and promotion of self-regulated learning strategies. The high percentages were corroborated by the respondents' comments. For example, one of the undergraduates wrote:

Generally, the tutors at the Writing Center have provided me with critical analysis that encouraged me to take more initiative to support my arguments, while also improving my writing self esteem in general. I know that without this service, my marks would have suffered immensely, and I would have experienced a much more frustrating academic year. Without this service I would have felt neglected by the University (S35).

Table 6. Statements Related to Users' Perceptions about How the Center Did or Did Not Help Them

No.	Statement	%
1	I have come to understand that writing is a process.	80
2	I have learned strategies for revising.	85
3	I have learned strategies that I can apply to future papers in other courses.	80
4	I have learned strategies to continue improving my oral communication skills on my own.	25
5	I have learned strategies to continue improving my writing on my own.	81
6	I have learned to teach myself to find and correct grammatical problems (self-editing and proofreading strategies).	74
7	I have learned to write with more clarity.	83
8	I have learned to use sources effectively and to cite them appropriately.	72
9	I have learned to organize my writing in a logical, coherent manner.	75
10	I have made progress in my thesis writing.	75
11	My ability to write has improved overall.	81
12	My confidence in my ability to complete written assignments has improved.	75
13	My ability to discuss issues in class has improved.	27
14	My ability to make presentations in class/conferences has improved.	25
15	My confidence about engaging in class discussion has improved.	24
16	My confidence about making presentations in class/conferences has improved.	30
17	The information in the workshop(s) has helped me to perform better in courses.	44
18	I feel more comfortable using the Writing Center as a learning resource.	81
19	When I sought help from the Writing Center, the grade I received for that assignment improved.	42
20	The Writing Center has helped me in my academic studies.	58

3. Satisfaction Survey

A total of 108 nonuser and 184 user respondents from undergraduate and graduate levels across disciplines completed a detailed satisfaction survey. Table 7 provides the demographic characteristics of respondents who completed the survey. Results from the survey showed that 32.6% visited the center once, 52.6% visited from two to five times, 10.5% visited from six to ten times, and 4.2 % visited more than ten times during the 2009–2010 academic year. Among the 108 respondents who indicated that they

Huang / Are We Having the Effect We Want?

had not used the center, 58% cited the reason that they did not need to use the services, 27.8% indicated that they did not know about the center's existence, 20.8% received help from others, and 28.4% reported that they were too busy to use the center. In terms of what services they used, 93.4% indicated those mainly related to writing support, with 60.4% using the one-on-one writing tutoring service and 33% using the drop-in service. Among users, 84.8 % indicated that they had not attended any workshops, and the remaining users had attended mainly writing-related workshops.

Table 7. Demographic Profile of Writing-Center Users (*N* = 184)

Gender	Female	71.7% (undergraduate); 61.2% (graduate)
	Male	27.7% (undergraduate); 37.3% (graduate)
Level of study	Undergraduate	70.7%
	Graduate	24.9%
	Diploma and other	4.4%
Year of Study	1	8.8% (undergraduate); 22.1% (graduate)
	2	36.4% (undergraduate); 23.5% (graduate)
	3	26.9% (undergraduate); 14.7% (graduate)
	4	28.5% (undergraduate); 4.4% (graduate)
	5 and above	9.3% (undergraduate); 35.3% (graduate)
Division	Humanities	24.7%
	Social Sciences	50.5%
	Physical Sciences	10.5%
	Life Sciences	13.2%
Language	EL1	83.3%
	EAL	16.6%; 29 different languages

Note. EL1 = English as a first language; EAL = English as an additional language. Percentages of users in each category do not add up to 100 because of cases in which respondents specified "other."

Results from the third section of the survey on support expected and received, as well as satisfaction levels, which are presented in the order of what users expected to receive from the most to the least, are summarized

in Table 8. The overall satisfaction rate was 77%, with work on organization, clarity, and conciseness receiving the top-three satisfaction ratings from users, and work on planning a paper, writing for a specific audience, and developing a format receiving the lowest satisfaction ratings. The overall satisfaction rate of 77.01% was tabulated on the basis of the satisfaction level in relation to individual items (see Table 8), which was corroborated by the result at the end of the survey, where users were asked to rate their overall satisfaction with the help they received; this rating was 76.1%.

Respondents were asked to explain the basis for their overall satisfaction rating. An analysis of the 92 responses that users provided generated the following key areas. Responses that were positive centered on three key areas: (a) tutors' traits (key descriptive words were "helpful," "respectful," "supportive," "patient," "professional," "positive," "friendly," and "encouraging"), (b) content (e.g., "tools," "strategies," and "explanation" of "APA rules," and "grammar rules" received), and (c) self-perceptions (e.g., confidence enhanced and writing improved). Responses that were linked to lower satisfaction centered mainly on four key areas: (a) tutors' traits (e.g., lacking the discipline-specific knowledge, language skills, and expertise required to help advanced writers), (b) content and delivery (e.g., workshop approach, focus too much on basic grammar), (c) outcomes (e.g., grades), and (d) time allotted per session (over 40% of respondents' comments were about this final area).

Beyond providing a measurement of users' satisfaction levels, the findings also, more importantly, generated data about the specific support that users most expected, and whether or not the support received was what they had expected. When the users were asked about whether they had informed the tutors they worked with about the specific help they needed, 91.6% indicated that they often or always clarified their need to tutors. When asked whether or not the tutor with whom they worked had clarified what he/she intended to focus on during the session, 61% of the users reported that their tutor often or always did.

Table 8. Tutoring Support Expected and Not Received, and Overall Satisfaction Level

Support	Expected	Not Received	Satisfaction Level*
Grammar	61%	15.1%	82.5%
Organization	58%	11.4%	90%
Clarity	54.1%	11.1%	87.8%

Huang / Are We Having the Effect We Want?

Support	Expected	Not Received	Satisfaction Level*
Conciseness	53.2%	14.3%	84.8%
Development of ideas	53.2%	13%	80%
Format	43.6%	20.3%	66.7%
Sentence structure	43.4%	9.8%	77.4%
Citation and documentation	43.3%	12.5%	73.3%
Building argument	42.2%	13.9%	71.8%
Thesis statement	40.3%	14%	75%
Punctuation	39.2%	10%	88.5%
Word choice	39%	13.3%	84.6%
Supporting ideas	38%	9.3%	81.8%
Logic	36.1%	15.8%	78.2%
Transitions	35.1%	12.3%	80.9%
Thesis or dissertation concerns	35.1%	13.8%	73.9%
Writing for a particular audience	34.7%	14.3%	66.6%
Paragraph writing	33.8%	17.2%	77.3%
Self-editing strategies	33.8%	20.7%	72.7%
Discipline-specific writing	33.8%	21.3%	71.4%
Summarizing	30.4%	14.8%	78.5%
Clarifying an assignment	28.8%	14.3%	75%
Planning a paper	26.8%	14.0%	56.3%
Paraphrasing	23.9%	16.4%	73.3%

Note. The satisfaction level here indicates the percentages of somewhat and very satisfied.

The fourth section of the survey concerns tutor and workshop satisfaction. The numeric results from this section on items related to tutors generally supported the qualitative data provided by respondents. It was understood that respondents might have worked with more than one tutor. In such cases, respondents were reminded to provide an overall rating that best represented their agreement with the statements presented in Table 9. The results demonstrated users' overall satisfaction with the tutors who had interacted or worked with them. When respondents were asked about whether they planned to use the Writing Center again, 60.6% strongly agreed with the statement, and 15.2% somewhat agreed. In terms of the

WPA 35.1 (Fall/Winter 2011)

level of satisfaction with the workshops they had attended, results are based on slightly over 15% of the respondents. Among the 14 workshops and workshop series offered during the 2009–2010 academic year, those that had the highest levels of satisfaction included Managing your Thesis or Dissertation for Graduate Students Series ($n = 9$), Writing Tips Series: Stronger Sentences ($n = 4$), and Writing Tips Series: Grammar ($n = 3$). When users were asked how well the tutors/instructors had helped them improve their oral communication skills, 21 of the 34 users (61.8%) responded indicated “good” or “excellent.”

Table 9. Level of Agreement in Response to Statements Concerning Tutors

Statement	Level of Agreement*
The tutor was friendly and made me feel welcome.	85.9%
The tutor worked with me in a respectful and professional manner.	91.3%
The tutor listened to me and took my work seriously.	89.7%
We dealt with the issues and concerns that I had in mind.	72.4%
The tutor helped me to think critically about my writing.	74.5%
The tutor communicated the information clearly and effectively.	89.5%
The tutor helped me to do my own work, rather than doing the work for me.	90.1%
The workshop presenter communicated the information clearly and effectively.	87.5%
I was able to apply what I learned in the session to future writing.	75%
The learning resources provided by the Center are useful.	81.5%
The tutor is knowledgeable about writing.	87%
I left knowing what I needed to work on to improve my writing.	77.6%

Note. The agreement level here indicates the percentages of somewhat and very strongly agree.

In the fifth section of the satisfaction survey, respondents were asked about their future involvement with the Writing Center. A total of 50.7%

of respondents indicated that they would use the tutoring services more if their schedules fit the center's hours. This total is followed by 57.9% who indicated that they would do so if they understood how the center's offerings could help them, and 43.6% who would if they understood how the tutors could help them. A total of 60% of users indicated that would use the tutoring services more if they needed more help, and 16.4% would attend if certain workshops were offered in different formats (e.g., webinars, one-on-one coaching), or with different content (e.g., discipline-specific writing, critical reading for academic purposes, planning a scientific paper, writing at the graduate level in specific disciplines, research-proposal workshops for sciences, workshops for advanced-level, competent writers and speakers). Respondents' repeated reference to and mention of support targeted to advanced or specialized writers or communicators in comments such as the following point to an area that the center can further develop: "The service is fine for undergraduates who need more general help with writing, but it didn't help me with my graduate-level, discipline-specific questions or writing. . . ." Another commented: "I'm a strong writer, and I wanted a set of critical eyes to go over my SSHRC applications [. . .]. I was a little frustrated by my tutor's [. . .] command of the English language, which made debates about word connotations and grammar a little challenging." Finally, when asked whether they would recommend the Writing Center to another person, 85.3% responded "yes" and 14.7% responded "no."

DISCUSSION AND IMPLICATIONS

The difficulty involved in assessing writing ability is well recognized in the field of language assessment, and variability in judgments of writing quality or ability is a specialized area of study in the language-testing field. To capture this difficulty, Gamaroff identified rater reliability in language assessment as "the bug of all bears" (31). Numerous studies of variability in assessments of writing performance have illuminated various possible sources and factors that may be connected to variability in the scoring methods used (e.g., analytical vs. holistic ratings), including those that are task-related (e.g., task types, topics, text types or genres, time), rater-related (e.g., language and cultural backgrounds, professional experience, age), and context-related (e.g., sequencing of the writing samples, writing/testing conditions) (e.g., Bachman, Lynch and Mason; Eckes; Johnson and Lim; Schaefer; Schoonen). Existing studies have produced mixed results. A few have reported high inter-rater reliability (e.g., Niller, *Number Speaks*), and this reporting might give practitioners a false impression that achieving acceptable inter-rater reliability is the norm rather than the exception; oth-

ers have identified substantial degrees of variability associated with raters' characteristics, types of tasks, scoring procedures, and contextual factors demonstrated in studies in the language-testing field (Erdosy).

Unlike existing studies that have achieved an acceptable degree of reliability through, for example, a necessary narrowing to a certain type of text from a particular course (e.g., Niller, *"The Number Speaks" Again*), this project's learning outcomes-based evaluation involved evaluating writing samples across disciplines with multiple text types and genres that most writing centers encounter when dealing with students across disciplines, and that are completed in non-testing, uncontrolled contexts. Even though (a) the time between the before- and after-drafts was controlled, (b) the rating procedures involved both analytical (i.e., following specific, identical scoring rubrics) and holistic ratings, and (c) the assessment involved three experienced raters, multiple sources of variance were anticipated, and they manifested in the degrees of correlation between and among raters. As Gamaroff and many others have pointed out, rater variation is expected, "owing to the fact that language is closely connected to human rationalities, imaginations, motivations and desires, which, because they each comprise an extremely complex network of biological, cognitive, cultural and educational factors, could easily compromise the quest for objectivity" (34). It is questionable whether there is such thing as a bias-free judgment and whether there can ever be a "true" rating. Raters may differ in the amount of importance they attach to various criteria. The same scores among the three raters do not necessarily indicate similar judgments, nor do different ratings among raters necessarily indicate different judgments (Gamaroff 32).

One might argue that such measures as limiting the number of courses where writing samples of only certain specific types are gathered, collecting more samples from the same courses across disciplines, adding more raters, or providing more training for raters in order to minimize variability in their judgments of writing samples would provide ratings with high inter-rater correlations. One could counter-argue, however, that, in a center designed to serve students across disciplines who have different writing demands, an assessment must entail the scope of the evaluation that was attempted in this project. There is also no guarantee that increased rater training would eliminate variations in ratings. Further, even extensive training and feedback regarding raters' ratings have been shown to be much less effective at reducing rater variability than researchers expected (e.g., Barrett; Elder, Knoch, Barkhuizen, and von Randow; Lumley and McNamara; Lumley; Weigle). Above all, how realistic is it for writing cen-

ters to allocate resources for more independent raters of hundreds of drafts, especially in times of financial difficulties and cutbacks?

Even though inter-rater reliability between and among raters is moderate to low, the resulting combined ratings, nonetheless, provide an important piece of the puzzle concerning overall performance in terms of each macro- and micro-level feature and as a whole. First, given the level of consistency within raters, a comparison of each individual rater's ratings of the before- and after-drafts and the trend lines of their ratings over time offers an important source of evidence for validating the impact of writing-center work. Mean ratings provided by each of the three individual raters were higher for *all* after-drafts than they were for before-drafts, with 11 out of 21 mean differences reaching the standard significance level (Table 2). Second, correlations between individual analytic features and the holistic rating exhibited similar association patterns; the similarity suggests that the three raters derived their individual judgments about the holistic ratings in a consistent way.⁵ Third, the inclusion of writing samples from a wide variety of disciplines, including multiple text types and genres, which is the daily reality of what most writing centers address, would make achieving high coefficients the exception rather than the norm. Even when high reliability is achieved, no one can guarantee that the same outcomes would be derived if the study were ever to be replicated. As such, criticisms concerning the potential lack of generalizability in learning outcomes-based evaluation may have gravely overlooked the comparison of each individual rater's ratings of the before- and after-drafts and the trend lines of their ratings over time (e.g., Tables 2, 3, and 4). One must note that a measure's reliability sets an upper limit on the strength of its correlation with other measures. If differences in means *over successive occasions* or in trend-line analyses (as was previously reported) emerge as statistically significant, however, the findings indicate that, even with the error variance, the effect was sufficiently strong to emerge as significant. In other words, noteworthy effects were found despite the reliability challenge.

In terms of the perception-based assessment, the average agreement concerning support in the speaking domain (e.g., academic oral conversation skills, oral presentation skills) (i.e., items 4, 13, 14, 15, and 16) provided by the center was 26.2%. This low level of agreement, in conjunction with results generated from the satisfaction survey and the usage profile analysis, suggest an opportunity to increase students' awareness of the range of services and support that the center provides. Concerning the statement regarding whether the support they received had helped them perform better in concrete terms (e.g., grades), 42% agreed. This percentage could be considered high, given that developing writing skills takes time and that the

survey was not conducted immediately after users' Writing Center visits, when, as anyone working in the writing-center field knows, the feedback is often overwhelmingly positive. Those who responded to the survey had time to see how they performed in the courses or assignments for which they had sought help or support from the center.

The perception-based survey might be considered limited by the low response rate (i.e., 6.9% out of the total number of different users of the center in fall 2009), which does not fall into the "magic sampling fraction" (74), according to Dörnyei, and because of the possibility that users who respond to a survey may be a particularly eager group of participants, rather than a cross-section (and, as such, the results can provide only some perceptions related to the center's impact on respondents' writing abilities or confidence levels). Still, the results are valuable in that, first, the relatively high percentage (78.5%) of respondents who agreed or strongly agreed about the center's positive impact in the areas presented in Table 6 indicates that the center is successfully helping students "develop [their] abilities and confidences as [writers]," according to the center's mission statement. In addition, information obtained about how users learned about the center, their reported main reasons for visiting the center, and low percentages in the speaking domain point to an opportunity to increase students' awareness of the range of services and support that the center provides. This recommendation is also supported by results from the satisfaction survey, where respondents indicated that they would use the center more if they understood how the center's offerings or tutors could help them.

Finally, results from the satisfaction survey provided indirect evidence about the value of the services and support provided by the center and its team of tutors. Such gathering of evidence serves several important purposes. First, this evidence indicates in a concrete way how those involved in writing-center work are contributing to students' educational experience. Second, the evidence provides a fuller depiction that can be used in assessing the center's services and support. Users' high level of satisfaction with the support received in organizing a paper, for example, was reflected in the significant improvement in the before- and after-drafts, as evaluated by all raters. These assessments of users' work provide a firmer ground for claiming the impact of the center's work. In addition to illuminating the center's strengths, results from the survey components, in combination with sources of data provided by other components utilized in this systematic, purposeful assessment project, serve as signposts that can direct efforts and limited resources toward best meeting the needs and expectations of the students that the center is designed to support. Further to the recommendations related to such matters as staffing, session length, and frequency of visits,

which were offered in an earlier report on the center's usage profile (Huang, *Forest of Forests*), findings reported in this paper suggest the need to consider or reconsider the following areas:

1. Reevaluate the length of tutoring sessions, a step supported by the usage-profile analysis reported in a previous paper, by the learning outcomes gleaned from the learning outcomes-based evaluation, and by the satisfaction survey.
2. Create a checklist of areas of focus, such as the one presented in Table 8, as the point of departure, where tutees and tutors can clarify their needs and expectations during the session. Recognize the importance of beginning tutoring at points where tutees perceive that support is needed and guiding tutees so that they can accurately self-diagnose areas where they need to devote effort and attention.
3. Raise students' awareness of what and how the center can support their academic studies. In addition to what the center is already doing (e.g., orientation events, class visits, connections with faculty members), consider creating a program that would draw on the volunteer services of faculty members who are willing to answer students' communication questions about writing or speaking in their disciplines, professions, or areas of expertise. This kind of work would strengthen the standing of writing-center work as part of the institution's learning and teaching fabric and also mediate the need for greater discipline-specific expertise in expressing ideas that convey users' specialized, sophisticated knowledge.
4. Design and implement writing and communication workshops that support the needs of students in specific disciplines, especially the physical and life sciences.
5. Conduct ongoing, systematic assessments that will strengthen the Writing Center's impact on students' learning and experience. Such assessments are needed to facilitate longer-term trend-line analyses that can help address demographic and enrollment trends at the institutional level.
6. Include the Writing Center as part of the student-experience survey conducted by the university. This inclusion would encourage an important shift toward both thinking about and viewing the Writing Center as an integral aspect of each student's entire educational experience.

The feasibility of an investigation that could be undertaken beyond the present academic institution was carefully considered. For those who have attempted or are contemplating an outcomes-based assessment in the field of writing-center research, the assessment methods employed in any outcomes-based assessment that goes beyond collecting usage data and constructing a usage profile, as suggested in another report (Huang, *Forest of Forests*), will hinge upon the real time and resource constraints that each writing center faces. Given today's budget pressures, the allocation of funds for such endeavors constitutes a challenge, but, without evidence obtained through systematic and/or longitudinal studies, administrators and practitioners cannot fully address the question: "Are we having the effect we want?" Both the process and product illuminated through this project point to the possibility for collaborative efforts by writing program administrators/writing-center practitioners and researchers in fields related to the work of writing centers when conducting such assessments.

LIMITATIONS AND FUTURE RESEARCH

The results must be considered within the context of the study's limitations. First, collecting authentic assignments produced by students for real course requirements as opposed to work done for research purposes necessarily involves evaluating students' writing with different lengths, purposes, tasks, and settings. Despite the control of a one-week period between the submissions of the before- and after-drafts, the myriad of individual and contextual variables and the individual raters' attributes unavoidably interject sources of variability in the rating process and outcomes. The validity and appropriateness of examining each individual rater's ratings have been clearly demonstrated, however.⁶ This analysis and the findings by no means suggest that one should forgo the attainment of higher inter-rater reliability. Collecting samples from the same variety of courses over multiple terms and using the same raters, who share similar rater characteristics, would facilitate a sub-level analysis by subject areas and possibly minimize potential sources of rating variability. These efforts would only further enhance the evidence provided by such assessment endeavors.

Second, the learning outcomes-based assessment focused on one-on-one writing support provided by the center. Gathering sources of direct evidence to validate the effectiveness of writing or speaking workshops offered by the center was beyond the scope of this assessment component. Instead, in the current project, workshops were assessed through the perceptions- and satisfaction-based components. With the knowledge that the validity of measuring learning outcomes on the basis of students' attendance at a

single 50-minute or 1.5 hour workshop may be limited, a follow-up study designed to focus on specific workshops would require different methods to assess the impact of such support on students' learning outcomes.

Third, the multi-component approach to writing-center assessment implemented in this project provided multiple sources of data to better understand the Writing Center's role in promoting students' learning and experience in relation to the support and services that the center has been set up to provide. Assessments of this scope and data of this kind are often lacking for various reasons, as pointed out previously in the paper. One should recognize that no single study can address all questions about the efficacy of a program or unit, nor can it fully measure its impact at all levels. This initiative, however, represents an important attempt to understand the overall functioning of a unit that has been established for just two years at the time this project commenced. The initiative is also consistent with the framework advanced by Betsy Barefoot in the context of the first-year seminar and also advocated by Lerner. A fuller understanding of any writing center's impact must be complemented by follow-up studies that, like the one presented here, involve multiple components. In addition, these studies need to use other methods, such as case studies of individual tutoring sessions and longitudinal studies that follow and track how groups of writing-center users develop their academic writing or speaking skills over the years.

CONCLUSION

Questions about what students gain from their participation in a particular academic program are at the heart of learning in higher education and, as such, an evidence-based approach to outcomes assessment deserves serious attention. In line with Lerner's vision that "mov[ing] toward 'measurable' outcomes can [. . .] be a potential opportunity rather than an impending threat" (2001: 4), this multi-component, outcomes-assessment study responds to the call for research in a specialized field that empirically substantiates the impact of writing-center work and can also be used to enhance the center's practice and operation. The study's findings present a complex picture of students' use of the Writing Center and the center's role in promoting the development of the academic English-language skills that it has been designed to support. The findings have immediate implications for recognizing the Writing Center's important work and for current practices at the institution where the assessment was conducted. Beyond those implications, this project provides practical guidance to writing program administrators and researchers seeking to conduct outcomes assessments. The description of the multi-component assessment method, which utilized

complementary sources of data, and its implementation should be useful to administrators, researchers, and practitioners in academic language-support units at many institutions of higher education. As institutions demand or undertake more writing-center assessments, the increased data and information will make it possible for writing centers to gauge their own effects, to make well-informed programmatic and pedagogical decisions, and to serve as guides in implementing changes that reflect their own institutional contexts.⁷

NOTES

1. The center usage-profile analysis was reported in an earlier report (Huang, *Forest of Forests*).

2. The ethics protocol for the project was approved by the university's Research Ethics Board in September 2009, with the Protocol Number 09–319.

3. Rater characteristics: All have advanced degrees in English, linguistics, applied linguistics, or education; one EL1 speaker and two EL1-like speakers; mean age: 50.3; mean years of teaching experience in academic writing: 19; and mean years of experience in rating: 8. All three raters received either formal or informal training in performance-based writing assessment; all have taught or are currently teaching at universities.

4. The same line of argument can be made about the variability involved in studies that used course grades from different courses taught by different instructors as sources of an indicator for comparing users and nonusers and/or pre- and post-center-visit assignment grades.

5. A qualitative post-rating response was elicited from each rater to understand his or her process of developing a holistic score for a piece of writing and to determine whether the rater employed a particular focus when forming a general impression of a draft's quality. All three raters, despite the variability among their ratings, reported placing a similar emphasis on the macro-level features as the primary focus that affected their impression of a draft's quality while deriving their holistic ratings. Each rater was provided a draft that had been assigned distinctively different ratings among the three raters. Each rater was then asked to "think aloud" as they re-rated the draft. Finally, raters were asked to reflect on whether they encountered any challenges in reconciling authentic subjectivity and objective precision as they followed the rubrics, and what they did when they found themselves having to reconcile their overall impression of a particularly hard-to-evaluate draft and the rating descriptors. The raters' responses revealed their individually complex and richly informative thinking and decision-making processes. Although rater judgment is not the focus of this paper, and adequate coverage of it would require more space and study, the greater understanding of

raters' thinking and decision-making processes obtained through this research has implications for future writing-center assessment endeavors.

6. A detailed qualitative study is currently underway that focuses on examining textual differences between the before- and after-drafts gathered from users who submitted three or more written assignments over the course of an academic term.

7. I owe a debt of gratitude to Catherine Mateer and Teresa Dawson for their vision and support that made this study possible; to Julie Bauer Morrison for her permission to adapt her survey instrument; to Lauren Charlton and Norman Thom, who assisted with distribution of the surveys; to Dahlia Beck, Brittney O'Neill, Chris Whitney, Shu-min Huang, and Carrie Hill for their assistance with data management and rating aspects of the project; and above all, to the Writing Center staff, the participants, and their inspiring instructors who supported this important initiative.

WORKS CITED

- Ady, Paul. "Fear and Trembling at the Center: Student Perceptions about the Tutorial." *The Writing Lab Newsletter* 12.8 (1988): 11–12. Web. 1 Sept. 2010.
- Allen, Mary J. *Assessing Academic Programs in Higher Education*. San Francisco: Jossey-Bass, 2004. Print.
- Bachman, Lyle F., Lynch, Brian K., and Mason Maureen. "Investigating Variability in Tasks and Rater Judgments in Performance Test of Foreign Language Speaking." *Language Testing* 12 (1995): 238–57. Web. 1 Sept. 2010.
- Barefoot, Betsy. "Evaluating the First-Year Seminar." Policy Center on the First Year of College. Web. 1 Sept. 2010.
- Barrett, Steven. "The Impact of Training on Rater Variability." *International Educational Journal* 2 (2001): 49–58. Web. 1 Sept. 2010.
- Bell, Elizabeth. "A Comparison of Attitudes Toward Writing." *The Writing Lab Newsletter* 7.2 (1982): 7–9. Web. 1 Sept. 2010.
- Bell, James H. "When Hard Questions Are Asked: Evaluating Writing Centers." *Writing Center Journal* 21.1 (2000): 7–28. Web. 1 Sept. 2010.
- Bennett, James. "The Effects of Instructional Methodology and Student Achievement Expectations on Writing Performance in Community College Composition Classes." Diss. University of Washington, 1988. Print.
- Carino, Peter, and Doug Enders. "Does Frequency of Visits to the Writing Center Increase Student Satisfaction? A Statistical Correlations Study—or Story." *The Writing Center Journal* 21.3 (2001): 83–103. Web. 1 Sept. 2010.
- Clark, Irene L. "Portfolio Evaluation, Collaboration, and Writing Centers." *College Composition and Communication* 44.4 (1993): 515–24. Web. 1 Sept. 2010.
- Cushman, Tara, Lindsey Marx, Carleigh Brower, Katie Holahan, and Elizabeth Boquet. "Using Focus Groups to Assess Writing Center Effectiveness." *The Writing Lab Newsletter* 29.7 (2005): 1–5. Web. 1 Sept. 2010.

- David, Caroll, and Thomas Bubloz. "Evaluating Students' Achievement in a Writing Center." *The Writing Lab Newsletter* 9.8 (1985): 10–14. Web. 1 Sept. 2010.
- Davis, Kevin. "Improving Students' Writing Ability: The Effects of a Writing Center." *The Writing Lab Newsletter* 12.10 (1988): 3–5. Web. 1 Sept. 2010.
- Denzin, Norman. "Strategies of Multiple Triangulation." *The Research Act*. Englewood Cliffs, NJ: Prentice Hall, 1989. 235–47. Print.
- Dörnyei, Zoltan. *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Mahwah, NJ: Lawrence Erlbaum, 2003. Print.
- Eckes, Thomas. "Rater Types in Writing Performance Assessments: A Classification Approach to Rater Variability." *Language Testing* 25.2 (2008): 155–85. Web. 1 Sept. 2010.
- Elder, Catherine, Knoch, Ute, Barkhuizen, Gary, and Janet von Randow. "Individual Feedback to Enhance Rater Training: Does it Work?" *Language Assessment Quarterly* 2 (2005): 175–96. Web. 1 Sept. 2010.
- Enders, Doug. "Assessing the Writing Center: A Qualitative Tale of a Quantitative Study." *The Writing Lab Newsletter* 29.10 (2005): 6–9. Web. 1 Sept. 2010.
- Erdosy, M. Usman. *Exploring Variability in Judging Writing Ability in a Second Language: A Study of Four Experienced Raters of ESL Compositions*. TOEFL Research Report 70. Princeton, NJ: Educational Testing Service, 2004. Web. 1 Sept. 2010.
- Gamaroff, Raphael. "Rater Reliability in Language Assessment: The Bug of All Bears." *System* 28 (2000): 31–53. Web. 1 Sept. 2010.
- Hawthorne, Joan. "Approaching Assessment As If It Matters." *The Writing Center Director's Resource Book*. Eds. Christina Murphy and Byron L. Stay. Mahwah, NJ: Lawrence Erlbaum, 2006. 237–48. Print.
- Henson, Roberta, and Sharon Stephenson. "Writing Consultations Can Effect Quantifiable Change: One Institution's Assessment." *The Writing Lab Newsletter* 33.9 (2009): 1–5. Web. 1 Sept. 2010.
- Huang, Li-Shih. *Graduate and Undergraduate Students' Academic Communication Needs Assessment*. The Learning and Teaching Centre, University of Victoria, 2008. Print.
- . "Seeing Eye to Eye? The Academic Writing Needs of Graduate and Undergraduate Students from Students' and Instructors' Perspectives." *Language Teaching Research* 14.4 (2010): 517–39. Web. 1 Sept. 2010.
- . *A Forest of Forests: Constructing a Centre Usage Profile as a Source of Outcomes Assessment*. The Learning and Teaching Centre, University of Victoria, 2010. Print.
- Hyland, Teresa, Grace Howell, and Zheng Zhang. *The Effectiveness of the Writing Proficiency Assessment (WPA) in Improving Student Writing Skills at Huron University College*. Toronto, Ontario: The Higher Education Quality Council of Ontario, 2010. Web. 1 Sept. 2010.
- Johanek, Cindy. *Composing Research: a Contextualist Paradigm for Rhetoric and Composition*. Logan: Utah State UP, 2000. Print.

- Johnson, Jeff S., and Gad S. Lim. "The Influence of Rater Language Background on Writing Performance Assessment." *Language Testing* 26.4 (2009): 485–505. Web. 1 Sept. 2010.
- Jones, Casey. "The Relationship Between Writing Centers and Improvement in Writing Ability: An Assessment of the Literature." *Education* 122.1 (2001): 3–20. Web. 1 Sept. 2010.
- Lerner, Neal. "Counting Beans and Making Beans Count." *The Writing Lab Newsletter* 22.1 (1997): 1–3. Web. 1 Sept. 2010.
- . "Choosing Beans Wisely." *The Writing Lab Newsletter* 26.1 (2001): 1–5. Web. 1 Sept. 2010.
- . "Writing Center Assessment: Searching for the 'Proof' of Our Effectiveness." *The Center Will Hold: Critical Perspectives on Writing Center Scholarship*. Eds. Michael A. Pemberton and Joyce Kinkead. Logan: Utah State UP, 2003. 58–73. Print.
- Lumley, Tom. "Assessment Criteria in a Large-scale Test: What Do They Really Mean to the Raters?" *Language Testing*, 19.3 (2002): 246–76. Web. 1 Sept. 2010.
- Lumley, Tom, and Tim F. McNamara. "Rater Characteristics and Rater Bias: Implications for Training." *Language Testing* 12 (1995): 54–71. Web. 1 Sept. 2010.
- Matthews, Mike. "Student Perceptions of Writing Center Personnel: Degrees of Separation." *English in Texas* 25.3 (1994): 5–7. Print.
- Morrison, Julie Bauer, and Jean-Paul Nadeau. "How Was Your Session at the Writing Center? Pre- and Post-grade Student Evaluations." *The Writing Center Journal* 23.2 (2003): 25–42. Web. 1 Sept. 2010.
- Naugie, Helen. "How Georgia Tech's Lab Prepares Students for the Georgia Mandated Proficiency Exam." *The Writing Lab Newsletter* 5.4 (1980): 5–6. Web. 1 Sept. 2010.
- Neuleib, Janice. "Proving We Did It." *The Writing Lab Newsletter* 4.7 (1980): 2–4. Web. 1 Sept. 2010.
- . "Evaluating a Writing Lab." *Tutoring Writing: A Sourcebook for Writing Labs*. Ed. Muriel Harris. Glenview, IL: Scott, Foresman, 1982. 227–32. Print.
- . "Research in the Writing Center: What to Do and Where to Go to Become Research Oriented." *The Writing Lab Newsletter* 9.4 (1984): 10–13. Web. 1 Sept. 2010.
- . "Evaluating Writing Centers: A Survey Report." *The Writing Lab Newsletter* 9.4 (1986): 1–5. Web. 1 Sept. 2010.
- Newmann, Stephen. "Demonstrating Effectiveness." *The Writing Lab Newsletter* 23.8 (1999): 8–9. Web. 1 Sept. 2010.
- Niller, Luke. "The Numbers Speak: A Pre-test of Writing Center Outcomes Using Statistical Analysis." *The Writing Lab Newsletter* 28.7 (2003): 6–9. Web. 1 Sept. 2010.
- . "'The Numbers Speak' Again: A Continued Statistical Analysis of Writing Center Outcomes." *The Writing Lab Newsletter* 29.5 (2005): 13–15. Web. 1 Sept. 2010.

- Pemberton, Michael A. "The Writing Lab Newsletter as History." *The Center Will Hold: Critical Perspectives on Writing Center Scholarship*. Eds. Michael A. Pemberton and Joyce Kinkead. Logan: Utah State UP, 2003. 21–35. Print.
- Pemberton, Michael A., and Joyce Kinkead, eds. *The Center Will Hold: Critical Perspectives on Writing Center Scholarship*. Logan: Utah State UP, 2003. Print.
- Roberts, David, H. "A Study of Writing Center Effectiveness." *The Writing Center Journal* 9.1 (1988): 53–60. Web. 1 Sept. 2010.
- Sandlon, John. "The Effects of A Skills Center upon the Writing Improvement of Freshman Composition Students." *The Writing Lab Newsletter* 5.3 (1980): 1–3. Web. 1 Sept. 2010.
- Schaefer, Edward. "Rater Bias Patterns in an EFL Writing Assessment." *Language Testing* 25.4 (2008): 465–93. Web. 1 Sept. 2010.
- Schoonen, Rob. "Generalizability of Writing Scores: An Application of Structural Equation Modeling." *Language Testing* 22.1 (2005): 1–30. Web. 1 Sept. 2010.
- Schuh, John, and Lee M. Upcraft. *Assessment Practice in Student Affairs: An Application Manual*. San Francisco: Jossey-Bass, 2001. Print.
- Sutton, Doris, and Daniel Arnold. "The Effects of Two Methods of Compensatory Freshman English." *Research in the Teaching of English* 8 (1974): 241–49. Web. 1 Sept. 2010.
- Tashakkori, Abbas, and Charles Teddlie. *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks: Sage, 1998. Print.
- . *Handbook of Mixed Methods in Social and Behavioural Research*. Thousand Oaks: Sage, 2003. Web. 1 Sept. 2010.
- Upcraft, M. Lee, and John Schuh. (2001). "Assessing the First-year Student Experience: A Framework." *Proving and Improving: Strategies for Assessing the First College Year*. Ed. Randy L. Swing. Columbia: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition, 2001. 7–9. Print.
- van Dam, Celine. "Effects of Writing Center Usage and Motivation on Academic Writing Performance." Diss. University of Southern California, 1985. Print.
- Waldo, Mark L. "More Than 'First-Aid': A Report on the Effectiveness of Writing Center Intervention in the Writing Process." *Issues in College Learning Centers* 5 (1987): 12–22. Web. 1 Sept. 2010.
- Weigle, Sara Cushing. "Investigating Rater/Prompt Interactions in Writing Assessment: Quantitative and Qualitative Approaches." *Assessing Writing* 6 (1999): 145–78. Web. 1 Sept. 2010.
- Wills, Linda B. "Competency Exam Performance and the Writing Lab." *The Writing Lab Newsletter* 8.10 (1984): 1–4. Web. 1 Sept. 2010.



APPENDIX A. SCORING RUBRICS

Score		Purpose	Development	Organization	Citation/ Format	Grammar	Punctuation
Superior (5)	A paper at this level to a high degree accomplishes all of the following.	Shows clear evidence of purpose, intention, and focus	Shows clear evidence that ideas/claims are extended, explained, elaborated upon, and clarified	Shows clear evidence of logically sequenced writing and coherent sequencing of ideas	Shows clear evidence that writing follows accurately and consistently the format required by the discipline	Demonstrates mastery of the standard conventions of grammar	Demonstrates mastery of standard conventions and mechanics
Strong (4)	A paper at this level to a high degree accomplishes all of the following.	Addresses the topic and task well, though may respond to some aspects of the task more effectively than others	Is generally well developed through the use of appropriate and sufficient explanations, examples, and/or details	Is generally well organized	Generally follows the format required by the discipline, with minor errors	Demonstrates strong control of the standard conventions of grammar; may have a few errors	Demonstrates strong control of standard conventions and mechanics; may have a few errors
Adequate (3)	A paper at this level has one or more of the following characteristics	Addresses only some part of the topic/task	Is adequately developed and generally supports ideas with reasons and examples	Is adequately organized, but the connection of ideas may be occasionally obscure	Adequately follows the format required by the discipline, with some errors	May have some errors	May have some errors

WPA 35.1 (Fall/Winter 2011)

Score		Purpose	Development	Organization	Citation/ Format	Grammar	Punctuation
Marginal (2)	A paper at this level may include one or more of the following weaknesses.	Is unclear or seriously limited in addressing the topic/task	Provides limited development in response to the topic	Is poorly organized	Has frequent errors in following the format required by the discipline	Has an accumulation of errors	Has an accumulation of errors
Weak (1)	A paper at this level is seriously flawed by one or more of the following weaknesses.	Lacks focus	Provides little development; simplistic development without support	Has very weak organization	Has serious and persistent errors in following the format required by the discipline	Is marred by persistent and serious errors	Is marred by persistent and serious errors
Incompetent (0)	Not connected to the topic, undeveloped, illogical, incoherent, disorganized; has serious mistakes, a format/citation structure that is not recognizable as following a particular style, and persistent errors in grammar and punctuation						