ASSESSING THE EFFECTS OF SERVICE-LEARNING ON STUDENT WRITING

Adrian J. Wurr
University of Arizona

This paper proposes methods to study the impact of service-learning on the writing performance of native and non-native English speaking students in first-year college composition. Linguistic and rhetorical features commonly identified as affecting judgments of writing quality will be compared to holistic essay and portfolio ratings to describe the impact of different teaching and learning contexts on writing performance. The implications of the study will be of particular interest to L1 and L2 university composition instructors interested in learning more about service-learning and writing assessment.

INTRODUCTION

Assessment in writing over the last several decades has been largely conducted through holistic ratings. Until recently, this has primarily involved the use of timed-essay tests for placement or diagnostic purposes. Portfolio assessment has been used with increasing frequency in recent years, primarily for summative evaluation purposes but occasionally for placement decisions too. However, these measures have been criticized by researchers as being insufficient for measuring specific strengths and weaknesses in student writing (Hamp-Lyons, 1996; Purves, 1985). Given that each assessment method has its niche--its strengths and weaknesses varying in respect to its purposes and context--a comprehensive writing assessment model should include what Elana Shohamy (1998) refers to as the "multiplism principle." That is, it should include multiple samples, measures, and methods to provide a broad, rich description of student writing performance. In consideration of the additional concerns for assessing students' writing in their second language, Liz Hamp-Lyons (1996) recommends that holistic essay evaluations be coupled with primary trait analysis--an alternate form of holistic assessment measuring the strength of specific linguistic and rhetorical features in a given text or writing sample--in order to gain a more complete profile of each student's writing ability.

This paper proposes a comprehensive writing assessment model to describe and measure the effects of a new instructional program featuring service-learning on the writing performance of first-year native and non-native English speaking college composition students. The model considers linguistic and rhetorical features in writing which, when compared to holistic evaluations of student writing and qualitative program assessments, will provide a more complete picture of the short-term impact of service-learning on student writing and learning. Due to space considerations, the focus of this paper will be on writing ability. Readers interested in the assessment of other learning outcomes are referred to a pilot study report (Wurr, 1999) on the subject.

RATIONALE

Service-Learning

Service-learning has attracted a great deal of attention over the last decade from educators, politicians, and community activists. According to one newspaper report (Waller, 1993), over

21% of all higher educational institutions had service-learning departments or offices in 1993. More recent and authoritative figures are currently being established by the National Service-Learning Clearinghouse (personal communication, April 23, 1999), but with increasing support from the government and private institutions, the numbers will almost certainly rise. Already in Arizona, several post-secondary educational facilities have recently adopted service-learning programs, including Arizona State University, the University of Arizona, and Chandler-Gilbert and Pima Community Colleges.

Brock Haussamen (1997), a service-learning coordinator at Raritan Community College, defines service-learning as "a new branch of experiential education" that combines traditional classroom learning with voluntary community service (p. 192). While experiential education only necessitates hands-on learning and active reflection, service-learning extends this to include social action in the form of participatory action-based research. "In a cycle of experience and reflection, students apply their skills and knowledge to help people, and in the classroom, they reflect on the people, social agencies, and communities they have encountered and on the nature of service" (Haussamen, 1997, p. 192). In emphasizing service-learning's theoretical home in experiential education, Haussamen and others (Giles & Eyler, 1994; Sheckely & Keeton, 1997; Cone & Harris, 1996) establish a strong foundation for the field in the related works of John Dewey, Kurt Lewin, Jean Piaget, and David Kolb. (e.g., Dewey, 1938; Kolb, 1984; Lewin, 1951; Piaget, 1977).

Composition specialists were among the first attracted to service-learning, based on the belief that students produce better writing when they are personally engaged in the writing topic (Cooper & Julier, 1995; B. Heifferon, personal communication, April 28, 1998). One of the first books in a planned series of eighteen volumes on service-learning in the disciplines published by the American Association for Higher Education was devoted to composition. *Writing the Community: Concepts and Models for Service-Learning in Composition* (Adler-Kassner, Crooks & Watters, 1997) presents many thoughtful chapters on composition courses using service-learning. The editors and Rosemary Arca (1997) discuss the beneficial impact service-learning can have on post-secondary basic writers, while Bruce Herzberg (1994/1997) presents a good discussion of the consciousness raising that students at a small liberal arts college experienced as a result of a year-long service-learning course cluster that coupled sociology and first-year composition with volunteer work as adult literacy tutors in an inner-city, halfway house. Useful and encouraging as these reports are, they can be faulted for a lack of scientific rigor, a point Adler-Kassner, Crooks, and Watters (1997) draw attention to in their summary of the research findings to date on service-learning and composition:

> Though the evidence is largely anecdotal, it points to a source in the sense that service-learning makes communication--the heart of composition--matter, in all its manifestations. Whether teaching, learning, planning and executing assignments, exploring the writing process, or even grading papers, students and instructors feel a greater sense of purpose and meaning in the belief that their work will have tangible results in the lives of others. (p. 2)

Missing from the AAHE collection is any discussion of using service-learning with non-native English speaking students. As Hamp-Lyons (1996) notes, there are at least two distinct

groups of ESL students on most American college campuses. One consists of immigrants who often have lived in the country for several years, attended American schools, and have a high degree of integrative motivation. Another consists of international students whose first day outside their home country is often their first day in an American classroom. Although well educated and highly motivated, international students do not usually intend to live in the United States permanently, and thus position themselves differently than do immigrant students in respect to the surrounding local community (Wurr, 1999). Such differences need to be accounted for and investigated more thoroughly in the service-learning and second language acquisition (SLA) literature. As Adler-Kassner, Crooks, & Watters (1997) note, educators need to gain "a better understanding of how ideologies connect and affect interactions and understanding" (p. 11) among all stakeholders in service-learning.

Although no published articles have investigated how students from diverse cultures and linguistic backgrounds might respond to service-learning, several informal reports[1] describing service-learning projects in ESL classes have noted generally positive learning outcomes. Noah Barfield (1999) provides a good overview of a service-learning unit he taught in a first-year composition course for both immigrant and international ESL students at Washington State University. The project involved students in researching environmental issues in an American city of their choice, analyzing the data from various perspectives, and then applying their knowledge to local projects such as writing information brochures for nonprofit agencies and cleaning up a local river bed. Though mostly descriptive in nature, Barfield's account claims an increase in student motivation, engagement, and writing quality.

Richard Seltzer (1998) has also involved lower-intermediate level ESL students at Glendale Community College in service-learning projects as conversation partners for senior citizens at a local nursing home. Satisfying the students' desire for native English conversation partners and the nursing home residents' desire for companionship, the project has been a win-win situation for everyone involved according to Seltzer.

Finally, the pilot study that I conducted (Wurr, 1999) investigated the impact of service-learning on native and non-native college composition students enrolled in English 102 and 108 respectively at the University of Arizona. Formal and informal writing assignments given before, during, and after the students' engagement in service-learning activities were analyzed to determine the effects of service-learning on students' writing, critical thinking, and perceptions of community, academia, and self. The results suggest that service-learning does appear to have a positive effect on participants' self-perception as students and community members, but that non-native English speaking students face greater challenges in successfully completing service-learning assignments than native English speaking students. The impact of service-learning on critical thinking and composing skills was less clear, though. This result, coupled with the dearth of empirical research on service-learning in composition and Second Language Acquisition (SLA), provides the impetus for the present study.

Assessment Procedures

The assessment model proposed here uses various data collection and analysis procedures to investigate the impact of service-learning on writing ability. Writing ability is operationalized

as a complex, non-linear, and dynamic system involving the interaction of several subskills and processes. This conceptualization of writing ability draws on the growing field of research known as chaos or complexity theory (see, for example, Larsen-Freeman, 1997; Gleick, 1987; Connor-Linton, 1995; Galloway, 1995; Bowers, 1990; Waldrop, 1992; Rogan, 1999; Wildner-Bassett, n.d.). Originating in the natural sciences, chaos theory attempts to describe natural phenomena in which the whole is greater than the sum of the parts.

Diane Larsen-Freeman, one of the first researchers to apply chaos theory to SLA, notes that language has many similarities to other dynamic nonlinear systems: It is complex, comprised of many subsystems such as syntax, morphology, phonology, lexicon, semantics, and pragmatics; and these subsystems are interdependent in that the strength of one is relative to the presence of others. "Thus, describing each subsystem tells us about the subsystems; it does not do justice to the whole of the language" (Larsen-Freeman, p. 149).

Chaos theory enhances the interpretation of writing samples in several ways. Primary trait scores—the assessed strength of a single trait or quality of writing such as syntax or coherence—may be compared to holistic scores, a numerical rating that describes the overall quality of a writing sample, to see the extent to which the parts, those individual characteristics of writing assessed through primary trait analyses, describe the whole. Further, chaos theory reminds researchers that "the whole" of writing performance and ability may be a larger and more complex phenomenon than the snapshots of a single writing sample or even of multiple samples collected over the course of a semester can adequately capture. Chaos theory encourages researchers to interpret results within a broader, more comprehensive frame of reference.

The writing assessment model proposed here uses a combination of holistic and primary trait assessments of writing samples to provide information on specific writing skills, and the interaction and application of those skills on specific writing tasks. This quantitative data will be combined with qualitative data on individual learner differences related to motivation and social orientation in order to gain a better understanding of the multiple factors at play in the study. This fits well with other models of service-learning evaluation proposed by Fenzel and Leary (1997) and Driscoll, Holland, Gelmon, & Kerrigan. (1996). According to these service-learning experts, a combination of qualitative and quantitative research techniques aids in measuring the impact of service-learning on all stakeholders—students, faculty, community and institution—and, it is argued here, better captures the dynamic interplay of elements within the teaching and learning context of the study.

CONTEXT

Recently there have been national (Office of Management and Budget, 1999) and local calls to make the research data produced at publicly funded Research I universities such as the University of Arizona more accessible to the general public. The Southwest Project is one local response to such concerns. Researchers, educators, and community activists in the local Tucson community are collaborating on an interdisciplinary, cross-cultural, and multi-institutional project to design instructional materials that are scalable to multiple audiences and purposes. Part of this effort involves students and teachers in several first-year composition courses at the University of Arizona collaborating with their counterparts in two local elementary schools to teach and

learn about the land and people of the Southwest. Native and non-native English speaking students in English 101 and 107 at the University of Arizona read and write about issues related to the Southwest in their college composition classes while also leading small group discussions on the Southwest in K-5th grade classes at Fort Lowell and Lawrence elementary schools.

The goals for English 101 and 107 as outlined in *A Student's Guide to First-Year Composition* (Wurr, Eröz, & Singh-Corcoran, 2000) are as follows. Students will:

1. Read texts to assess how writers achieve their purposes with their intended audiences.
2. Learn the conventions of scholarly research, analysis, and documentation.
3. Learn other conventions of academic writing, including how to write clear and correct prose.
4. Learn to revise and respond to feedback from readers to improve and develop drafts.
5. Learn to develop ideas with observations and reflections on [their] experience.
6. Learn to analyze and write for various rhetorical situations.
7. Develop a persuasive argument and support it with evidence and effective appeals that target [their] intended audience.

To demonstrate the degree to which these goals are met, participating students in selected sections write a personal experience, a rhetorical analysis, and a persuasive essay as outlined in Table 1.

Table 1. English 101/107 Southwest Project Essay Assignment Sequence and Descriptors

1. Rhetorical Analysis essay (5-7 pages), in which the students research a local environmental problem from various viewpoints.
2. Persuasive Essay (4-6 pages), in which students suggest ways to solve or reduce the impact of the environmental problem they researched.
3. Reflective Essay (4-6 pages) which will serve as a preface to a portfolio on students' accomplishments over the semester, and within which the students will explain why they chose the texts they did, whom they are intended for, and what purpose the texts or portfolio is meant to serve.

PURPOSE

The purpose of this segment of the study is to investigate valid and reliable methods for describing writing quality based on current linguistic and rhetorical theories for analyzing student writing, with particular regard to persuasive essay writing. With this in mind, the following research questions were posed:

1. What valid and reliable indicators of writing quality can be identified?
2. What is the relationship between the quantity of rhetorical appeals and essay quality?
3. What is the relationship between the quality of coherence and essay quality?
4. What is the relationship between characteristics of syntax usage and essay quality?
5. What is the relationship between characteristics of reasoning and essay quality?

MAJOR VARIABLES

While each educational context has unique characteristics of its own, some variables affecting judgments of writing quality tend to recur in many studies (e.g., Bamberg, 1983; Biber, 1986; Connor, 1990, 1995; Connor & Lauer, 1985; Lloyd-Jones, 1975). Moreover, statistical procedures such as Rasch measurement and Multiple Regression Analysis (MRA) have allowed researchers to account for the contribution of these variables along with the differing effects of a given writing prompt, scoring guide, and/or inter-rater variation on holistic writing assessments. For example, Hamp-Lyons and Henning (1991) were able to assess with reasonable accuracy the writing performance of adult non-native English speakers on seven major variables using holistic and multi-trait assessment tools combined with Rasch analyses designed for different educational contexts. Their study suggests that researchers do not have to design new instruments from scratch every time they want to assess writing in a new situation. Rather, with reasonable care and consideration, they may fine tune established reliable techniques to fit the local context. With this in mind, primary trait analyses that have reliably measured rhetorical appeals, coherence, syntax, and reasoning in other contexts are presented below along with impressionistic scoring mechanisms as suggested means for documenting the impact of service-learning on student writing performance.

Because holistic essay evaluation is a cognitively demanding task, there is a tendency for raters to gravitate towards the center when asked to assign separate scores to multiple-traits in a single essay reading. Thorndike and Hagen (1969) attributed this to a "halo" or carryover effect of one trait upon the other. Their recommendation to conduct separate readings with different ratings for each trait to be assessed will be followed in the present study.

Analysis of Rhetorical Appeals

Ulla Connor and Janice Lauer (1985) developed scales for judging the persuasiveness of student writing for use in the International Study of Written Composition (commonly referred to as the IEA study because of its sponsor, the International Association for the Evaluation of Educational Achievement) conducted by Alan Purves (1988) and other researchers around the world. Starting with the use of ethos, pathos, and logos as persuasive appeals first identified in Aristotle's *Rhetoric*, and integrating the work of more modern rhetoricians such as James Kinneavy (1971) and Lauer, Montague, Lunsford, and Emig (1985), Connor and Lauer (1985) describe measures for identifying and rating the use of three persuasive appeals: Rational, credibility, and affective. Outlined in Table 3, the rational, credibility, and affective appeal scales had interrater reliabilities of .90, .73, and .72 respectively in the IEA study (Connor, 1990, p. 76).

Table 3: Rhetorical Appeals Scale

Rational

0	No use of the rational appeal.*
1	Use of some rational appeals, minimally developed or use of some inappropriate (in terms of major point) rational appeals.
2	Use of a single rational appeal* or series of rational appeals* with at least two points of development.
3	Exceptionally well developed and appropriate single extended rational appeal* or a coherent set of rational appeals.*
	*Rational appeals were categorized as quasi-logical, realistic structure, example, analog.

Credibility

0	No use of credibility appeals
1	No writer credibility but some awareness of audience's values; or Some writer credibility (other than general knowledge) but no awareness of audience's values.
2	Some writer credibility (other than general knowledge) and some awareness of audience's values.
3	Strong writer credibility (personal experience) and sensitivity to audience's values (specific audience for the solution).

Affective

0	No use of the affective appeal.
1	Minimal use of concreteness or charged language.
2	Adequate use of either picture, charged language, or metaphor to evoke emotion.
3	Strong use of either picture, charged language, or metaphor to evoke emotion.

Note. From "Cross-Cultural variation in persuasive student writing," by U. Connor & J. Lauer, 1988, *Writing Across Languages and Cultures*, edited by Alan C. Purves, p. 138. Copyright 1988 by Sage Publications. Reprinted with permission.

Analysis of Coherence

Although teachers and researchers have identified coherence as an important aspect in the quality of written texts, defining exactly what is meant by coherence has proved to be a difficult task. The prevailing opinion seems to emphasize the interactions between the reader and the text in defining coherence. Phelps (1985) for example, defines coherence as "the experience of meaningfulness correlated with successful integration during reading, which the reader projects back into the text as a quality of wholeness in its meanings" (p. 21). But even she admits that definitions of "successful integration" may vary from one reader or rater to the next.

Research indicates that topical structure can be an important indicator of overall writing quality (Witte, 1983a, 1983b; Connor, 1990; Connor & Farmer, 1990; Cerniglia, Medsker, & Connor, 1990). Witte (1983b) found that high quality essays had more parallel and extended parallel progression than low quality essays. Simply put, good writers tend to elaborate on important ideas while weaker writers often stray from the point by introducing new ideas not relevant to the discourse.

Building on this idea, Bamberg (1983, 1984) developed a system to help students revise their essays and improve coherence using topical structure analysis. Connor & Farmer (1990) adapted this into a four-point rubric to measure text cohesion. Students responded positively and made significant revisions to early drafts of their essays using topical analysis (Connor, 1996, p. 87), while the researchers using the rubric achieved an interrater reliability of .93 (Connor &

Lauer, 1985, p. 311). Given such high interrater reliability, and in consideration of the fact that Bamberg's system for analyzing text cohesion has withstood "the test of peer review" (Connor & Lauer, 1985, p. 311), her system was chosen for the present study and is shown in Table 4.

Table 4. Bamberg's "FourPoint Holistic Coherence Rubric

The writer...		
	4	identifies the topic and does not shift or digress.
		orients the reader by describing the context or situation.
		organizes details according to a discernible plan that is sustained throughout the essay.
		skillfully uses cohesive ties (lexical cohesion, conjunction, reference, etc.) to link sentences and/or paragraphs.
		often concludes with a statement that gives the reader a definite sense of closure.
		makes few or no grammatical and/or mechanical errors that interrupt the discourse flow or reading process.
	3	meets enough of the criteria above so that a reader could make at least partial integration of the text.
	2	does not identify the topic and inference would be unlikely.
		shifts topic or digresses frequently.
		assumes reader shares his/her context and provides little or no orientation.
		has no organizational plan in most of the text and frequently relies on listing.
		uses few cohesive ties (lexical, conjunction, reference, etc.) to link sentences and/or paragraphs.
		makes numerous mechanical and or grammatical errors, resulting in interruption of the reading process and a rough or irregular discourse flow.
	1	essay is literally incomprehensible because missing or misleading cues prevented readers from making sense of the text.

Note. From "Understanding persuasive essay writing: Linguistic/Rhetorical approach" by U. Connor and L. Lauer, *Text, 5* (4), p. 311. Copyright 1985 Mouton Publishers. Reprinted with permission.

Analysis of Syntactic Features

The T unit, the smallest part of a sentence that can be considered a complete thought, has been the means of choice for analyzing syntactic patterns in student writing for the better part of the last 30 years in composition studies (see, for example, Hunt, 1965; Mellon, 1969; O'Hare, 1973; Stotsky, 1975). Yet computer technologies have more recently enabled researchers to analyze more complex syntactic patterns in writing. Douglas Biber (1985, 1986, 1987) has developed a multi-feature/multi-dimensional computerized method to explain over 120 linguistic variations commonly found in texts. Factor analysis was used to find group features that had high co-occurance rates. These were then described as textual features.

Two primary features identified by Biber are "interactive versus edited text" and "abstract versus situated style." Both exist as continuums rather than strict dichotomies. The interactive vs. edited text distinction contrasts features showing high personal involvement with those which allow editing and lexical choice. As is summarized in Table 5, *That* clauses, first person pronouns, second person pronouns, contractions, and the pronoun *it* were all associated with a high degree of interaction. Nominalizations, prepositions, passives, and specific conjunctions were features describing the abstract versus situated style continuum, according to Biber (1986).

Table 5. Selected Syntactic Variations

Interactive vs. Edited	Abstract vs. Situated Style
that clauses	nominalizations
first person pronouns	prepositions
second person pronouns	passives
contractions	specific conjunctions
pronoun *it*	

Note. From "Linguistic/Rhetorical measures for international persuasive student writing," by U. Connor, 1990, *Research in the Teaching of English, 24*(1), p. 71. Copyright 1990 by the National Council of Teachers of English. Adapted and reprinted with permission.

With the aid of a computer, the number of occurrences of each in a text can be counted, taking the sum of the totals for interactive vs. edited text features and abstract vs. situated style features to create factor scores for each essay.

Analysis of Reasoning

Toulmin (1958) presents a model of informal logic to "assess the soundness, strength, and conclusiveness of arguments" (p. 1) that is comprised of three main parts: claims, data, and warrants. Claims are defined as "conclusions whose merits we are seeking to establish" (p. 97). Data provides support for the claims in the form of experience, facts, statistics, or events. Warrants are "rules, principles, [or] inference-licenses" that "act as bridges" between claims and data (p.98). Connor & Lauer (1988) developed a three-point analytic scale to rate the quality of reasoning in persuasive essays using Toulmin's categories of claim, data, and warrant. Shown in Table 6, Connor and Lauer's scale assesses both the quality and the quantity of the logic used.

Table 6. *Criteria for Judging the Quality of Claim, Data, and Warrant*

Claim	1.	No specific problem stated and/or no consistent point of view. May have one subclaim. No solution offered, or if offered nonfeasible, unoriginal, and inconsistent with claim.
	2.	Specific, explicitly stated the problem. Somewhat consistent point of view. Relevant to the task. Has two or more subclaims that have been developed. Solution offered with some feasibility, original, and consistent with major claim.
	3.	Specific, explicitly stated problem with consistent point of view. Several well-developed subclaims, explicitly tied to the original major claim. Highly relevant to the task. Solution offered that is feasible, original, and consistent with major claims.
Data	1.	Minimal use of data. Data of the "everyone knows" type, with little reliance on personal experience or authority. Not directly related to major claim.
	2.	Some use of data with reliance on personal experience or authority. Some variety in use of data. Data generally related to major claim.
	3.	Extensive use of specific, well-developed data of a variety of types. Data explicitly connected to major claim.
Warrant	1.	Minimal use of warrants. Warrants only minimally reliable and relevant to the case. Warrants may include logical fallacies.
	2.	Some use of warrants. Though warrants allow the writer to make the bridge between data and claim, some distortion and informal fallacies are evident.
	3.	Extensive use of warrants. Reliable and trustworthy allowing rater to accept the bridge from data to claim. Slightly relevant. Evidence of some backing.

Note. From "Cross-Cultural variation in persuasive student writing," by U. Connor & J. Lauer, 1988, *Writing Across Languages and Cultures*, edited by Alan C. Purves , p. 145. Copyright 1988 by Sage Publications. Reprinted with permission.

The preceding methods of identifying and analyzing linguistic and rhetorical components of writing can provide a detailed profile of a student's writing performance on a persuasive essay assignment. As Hamp-Lyons (1996) notes, such multi-trait analyses can be especially beneficial in the case of assessing second language writers because it can balance language control with other salient traits of the writer's text. However, as was noted earlier, to better understand how these variables work together within a given text and context, an impressionistic score of the whole essay is also necessary.

Holistic Writing Assessments

In the book *Measuring Growth in Writing*, which many claim helped turn the tide against indirect measures of writing towards more valid holistic assessment procedures, Paul Diederich (1974) asserts that interrater reliability scores of .80 should be the minimum acceptable standard for program evaluation purposes. Countless large and small scale essay rating sessions since then have confirmed that such standards are easily attainable when raters from similar backgrounds are trained in the use of a scoring guide. However, Connor (1990) points out that such training on specific points in the prompt or text could confound the correlation between the independent variables identified in a study on writing ability and the holistic scores given to sample papers. Using a five-point impressionistic holistic scoring procedure, she was able to achieve interrater reliability rates slightly above the minimums set by Diederich (1974) for program evaluation purposes.

Impressionistic assessments of writing provide some advantages over primary trait scoring. In addition to accounting for the interaction of elements within a text, impressionistic scoring also allows for a greater degree of interaction between the reader, writer, and text than evaluations based on the enumeration of linguistic and rhetorical features in a text. Also, since the weight of any one element within a text is always relative to other factors, holistic assessments are less likely to penalize second language writers for surface level errors than primary trait scales concerned with accuracy and mechanics[2]. Students will have had sufficient time to revise and edit all writing samples submitted for evaluation. Hamp-Lyons (1996) notes that this reduces the likelihood of fossilized errors appearing, as they often do in timed essay writing, by allowing students to avail themselves of all available resources—including peer tutors, writing center consultants, the teacher, and computer grammar and spell check programs—before they submit their writing for evaluation. This helps ensure that the writing sample represents the student's true writing ability for the task rather than one artificially induced by a timed-essay test.

Although each teacher and class participating in the study would be encouraged to develop their own scoring guide or grading rubric that met program goals for each essay assignment, for research purposes all essays would be rated by a team of qualified independent raters using a five-point scale based on the grading criteria outlined in *A Student's Guide to First-Year Composition* (Wurr, Eröz, & Singh-Corcoran, 2000) and presented in Table 7. Since the independent raters in the present study will all be graduate teaching assistants in the Composition program at the University of Arizona, and thus familiar with course goals and scoring guide outlined above, a brief review of the scoring guide and sample essays should be

enough to achieve interrater reliability rates of at least .80 without compromising correlational data between holistic scores and the independent variables.

Table 7. Holistic Scoring Guide for Persuasive Essays

Score of 5: Excellent	Strong, clear focus and thesis. Effective organization—including a beginning, middle, and end—with logical grouping of ideas into paragraphs. Lots of details and relevant examples from outside sources and, when appropriate, personal experience to support main ideas. Discussion shows a clear understanding of issue and texts, as well as a sense of purpose and audience. Few errors.
Score of 4: Good	Clear focus and thesis. Overall coherence with paragraphs to group similar ideas. Some examples and supporting details. Discussion demonstrates a good understanding of the issue and integrates ideas from primary and secondary sources of information. Occasional errors.
Score of 3: Adequate	Weak focus and thesis. Some coherence and logical grouping of ideas. Some examples and details, though connections may not always be clear. Discussion demonstrates a basic understanding of the issue and texts. Multiple errors.
Score of 2: Poor	No clear focus or message. Few appropriate examples or details. Discussion relies on a limited number of sources of information and overlooks complicating evidence. Serious errors which interfere with meaning.
Score of 1: Failing	Writing is seriously incomplete or does not address the assignment prompt. Errors prevent communication.

Portfolio Assessment

Portfolios have become increasingly popular in the last decade as a means for assessing both L1 and L2 writing because, as White (1995) explains, "multiple measures are always better than single measures" (p. 38). Hamp-Lyons (1996) also notes that portfolios have ecological validity, also known as beneficial backwash, for second language writers because they provide opportunities for teachers, students, and other stakeholders to discuss the writer's individual strengths and weaknesses, as well as growth over time. For ESL students educated in American schools, such discussions can prod students to invest more time and energy in the writing process because, with more support and resources available to students during the composing process, higher standards for achievement are more readily accepted and attained (Hamp-Lyons, 1996, p. 237). Portfolios also provide greater contextualization of writing processes and products, allowing them to serve multiple purposes and audiences. Because writing in service-learning courses often serves different purposes and audiences than writing in traditional composition courses, portfolios were the best way to contextualize the writing students engaged in over the course of the semester.

For the purposes of this study, a writing portfolio is defined as a collaborative effort between students and teachers in which students present their accomplishments over the course of the semester. Students are allowed to choose what represents their best work, and both explain their choices and assess the outcomes in a reflective essay introducing their portfolio to the reader. Although allowing students to choose portfolio contents makes it more difficult to compare equivalent writing tasks and genres amongst all participants, the students' ability to "analyze critically and write for various rhetorical situations" is a major goal of the course and hence an appropriate part of assessment. The portfolio scoring guide in Table 8, originally created by Donald Daiker for use at Miami University, was chosen as providing a suitable

baseline for holistically evaluating portfolios while also emphasizing context, creativity, and risk-taking in each category descriptor.

Table 8. Scoring Guide for Portfolios

Score of 6: Excellent	A portfolio that is excellent in overall quality. It is characteristically substantial in content (both length and development) and mature in style. It demonstrates an ability to handle varied prose tasks successfully and to use language creatively and effectively. Voice tends to be strong, and there is a clear sense of audience and context. Often, there is a close connection between the writer's sense of self and the writing—and/or a sense of thematic unity within the different portfolio pieces. A "6" portfolio typically takes risks that work—either in content or form—and challenges the reader by trying something new.
Score of 5: Very Good	A portfolio that is very good in overall quality. It suggests the excellence that the "6" portfolio demonstrates. Typically, a "5" portfolio is substantial in content, although its pieces are not as fully developed as a "6", and it uses language effectively but not as creatively as a "6". It suggests an ability to handle varied prose tasks successfully, and its voice is clear and distinct if not powerful. Sense of audience and context is clearly present if not always firm. A "5" portfolio tends not to take as many risks as a "6."
Score of 4: Good	A portfolio that is good in overall quality. The writing is competent both in content and style. There are more strengths than weaknesses, but there may be an unevenness of quality or underdevelopment in one or two pieces. The reader may want "more" to be fully convinced of the writer's ability to handle varied prose tasks successfully and to use language effectively.
Score of 3: Fair	A portfolio that is fair in overall quality. It suggests the competence that a "4" portfolio demonstrates. Strengths and weaknesses tend to be evenly balanced—either within or among the four pieces. One or more of the pieces may be too brief or underdeveloped. There is some evidence of the writer's ability to handle varied prose tasks successfully and to use language effectively, but it is often offset by recurring problems in either or both content and style. A "3" portfolio often lacks both a clear sense of audience and a distinctive voice.
Score of 2: Below Average	A portfolio that is below average in overall quality. It does not suggest the writing competence that a "3" portfolio does. Weaknesses clearly predominate over strengths. The writing may be clear, focused, and error-free, but it is usually more thin in substance and undistinguished in style. Several pieces may be either short or underdeveloped or abstract and vague. Moreover, the writer rarely takes risks, relying instead on formulas and clichès. There is little evidence of the writer's ability to handle varied prose tasks successfully. The few strengths of a "2" are more than overbalanced by significant weaknesses.
Score of 1: Poor	A portfolio that is poor in overall quality. There are major weaknesses and few, if any, strengths. A "1" portfolio lacks the redeeming qualities of a "2." It is usually characterized by pieces that are unoriginal and uncreative in content and style. The portfolio seems to have been put together with very little time and thought.

Note. From Daiker (1992-3) as reprinted in *Teaching and Assessing Writing* (pp. 301-303), by E. M. White, 1998, ME: Calendar Islands Publishers. Adapted with permission of the author.

STATISTICAL ANALYSIS

The extent to which the major variables identified above as contributing to the students' overall writing quality will be determined by comparing primary trait scores to impressionistic scores. Using multiple regression analysis, Connor (1990), for example, was able to account for 61% of the variation between impressionistic ratings and 11 independent variables, including those proposed in the present study. Though there are significant differences in how Connor and I have approached the analysis of some variables, multiple regression analysis procedures can be applied in both cases to statistically describe the relationship between each of the major variables—rhetorical appeals, coherence, syntax, and reasoning—identified as contributing to writing quality.

CONCLUSION

Although the proposed assessment instrument is summative in design, the results from the study could usefully be applied to future curriculum design and instruction. Composition students and teachers would benefit from gaining a more informed understanding of the most salient writing traits in holistic judgments of writing quality, while those interested in service-learning would gain empirical support for their practices. Both would gain a greater understanding of how different student populations might respond to service-learning initiatives in college composition. All readers can benefit from the multiple perspectives provided by the various participants and stakeholders, as well as from the interdisciplinary nature of the proposed research design.

NOTES

4. The TESOL 2000 convention held in Vancouver, B.C. March 13-18 featured half a dozen or more presentations related to service-learning in ESL classes. Since these presentations occurred while this article was going to press, I have not been able to incorporate summaries of the work in this paper. However, I think the existence of so many presentations on service-learning at a international convention like TESOL is significant in that it shows an increased awareness of service-learning amongst TESOL professionals.
5. See White & Polin (1986), though, for another possible outcome of holistically scored ESL texts.

REFERENCES

Adler-Kassner, L., Crooks, R., and Watters, A. (1997). Service-learning and composition at the crossroads. In L. Adler-Kassner, R. Crook, and A. Watters (Eds.), *Writing the community: concepts and models for service-learning in composition* (pp. 1-18). Urbana: IL: NCTE.

Arca, R. (1997). Systems thinking, symbiosis, and service: The road to authority for basic writers. In L. Adler-Kassner, R. Crooks, and A. Watters (Eds.), *Writing the community: concepts and models for service-learning in composition* (pp. 133-142). Urbana: IL: NCTE.

Bamberg, B. (1983). What makes a text coherent? *College Composition and Communication*, *34*, 417-429.

Bamberg, B. (1984). Assessing coherence: A reanalysis of essays written for the National Assessment of Educational Progress, 1969-1979. *Research in the Teaching of English, 18* (3), 305-319.

Barfield, N. (1999). Service learning and ESL. [On-line]. Available: http://www.wsu/~noahb/slesl_overview.html.

Biber, D. (1985). Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses. *Linguistics, 23*, 337-360.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language, 62*, 99-119.

Biber, D. (1987). A textual comparison of British and American writing. *American Speech, 62*, 99-119.

Bowers, R. (1990). Mountains are not cones: What can we learn from chaos? In J. E. Alatis (Ed.), *Georgetown University round table on languages and linguistics* (pp. 123-136). Washington, DC: Georgetown University Press.

Cerniglia, C,. Medsker, K., & Connor, U. (1990). Improving coherence using computer-assisted instruction. In U. Connor & A.M. Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives*. Arlington, VA: TESOL.

Cone, D., & Harris, S. (1996). Service-learning practice: Developing a theoretical framework. *Michigan Journal of Community Service Learning, 3*, 31-43.

Connor, U. (1990). Linguistic/Rhetorical measures for international persuasive student writing. *Research in the Teaching of English, 24*(1), 67-87.

Connor, U. (1995). Examining syntactic variation across three English-speaking nationalities through a multifeature/multidimensional approach. In D. L. Rubin (Ed.), *Composing social identity in written language* (pp. 75-87). Hillsdale, NJ: L. Erlbaum Associates.

Connor, U. (1996). *Contrastive Rhetoric: Cross-cultural aspects of second-language writing*. NY: Cambridge University Press.

Conner, U., & Farmer, M. (1990). The teaching of topical structure analysis as a revision strategy for ESL writers. In B. Kroll (Ed.), *Current issues in second language writing research* (pp. 126-139). NY: Cambridge University Press.

Conner, U. & Lauer, L. (1985). Understanding persuasive essay writing: Linguistic/Rhetorical approach. *Text, 5* (4), 309-326.

Conner, U. & Lauer, L. (1988). Cross-cultural variation in persuasive student writing. In A. C. Purves (Ed.), *Writing across languages and cultures*, (pp. 138-159). Beverly Hills, CA: Sage Publications.

Connor-Linton, J. (1995). Complexity, linguistics, and language teaching. In J. E. Alatis, C. A. Straehle, B. Gallenberger, & M. Ronkin (Eds.), *Georgetown University round table on languages and linguistics* (pp. 596-604). Washington, DC: Georgetown University Press.

Cooper, D. C., & Julier, L. (1995). Writing the ties that bind: Service-Learning in the writing classroom. *Michigan Journal of Community Service Learning, 2*, 72-85.

Dewey, J. (1938). *Education and experience*. New York: Macmillan.

Diederich, P. B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Driscoll, A., Holland, B., Gelmon, S., & Kerrigan, S. (1996). An assessment model for service-learning: Comprehensive case studies of the impact on faculty, students, community and institution. *Michigan Journal of Community Service Learning, 3*, 66-71.

Fenzel, L.M. and Leary, T.P. (1997). Evaluating outcomes of service-learning courses at a parochial school. Paper presented at the Annual Meeting of the American Educational Research Association in Chicago, IL March 24-28. ERIC ED 410 281.

Galloway, V. (1995). From control to chaos: Reflections on an "unstylish" research paradigm. In M.A. Haggerstrom, L.Z. Morgan, & J.A. Wieczorek (Eds.), *The foreign language classroom: Bridging theory and practice* (pp. 123-136). NY: Garland Publishing.

Giles, D.E., & Eyler, J. (1994). The theoretical roots of service-learning in John Dewey. *Michigan Journal of Community Service-Learning, 1*(1), 77-85.

Gleick, J. (1987). *Chaos: A new science*. NY: Viking.

Larsen-Freeman, D. (1997). Chaos/Complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165.

Hamp-Lyons, L. (1996). The challenges of second-language writing assessment. In White, E. M., Lutz, W. D., & Kamushikiri, S. (Eds.). *Assessment of writing: Politics, policies, practices* (pp. 226-240). NY: Modern Language Association.

Hamp-Lyons, L., & Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple-trait scoring instrument across ESL writing assessment contexts. *Language Learning, 41*(3), 337-373.

Haussamen, B. (1997). Service-learning and First-Year Composition. *Teaching English in the Two-Year College.* October, 192-198.

Herzberg, B. (1994/1997). Community service and critical teaching. In L. Adler-Kassner, R. Crooks, and A. Watters (Eds.), *Writing the community: concepts and models for service-learning in composition* (pp. 57-70). Urbana: IL: NCTE. (Reprinted from Community service and critical teaching. *College Composition and Communication, 45,* 307-319).

Hunt, K. W. (1965). Grammatical structures written at three grade levels. (NCTE Research Report No. 3). Urbana: IL: NCTE

Kinneavy, J. L. (1971). *A theory of discourse.* Englewood Cliffs, NJ: Prentice-Hall.

Lauer, J. M., Montague, G., Lunsford, A., & Emig, J. (1985). *Four worlds of writing* (2nd ed.). NY: Harper and Row.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice-Hall.

Lewin, K. (1951). *Field theory in social sciences.* New York: Harper & Row.

Llyod-Jones, R. (1975). Primary trait scoring. In C. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 33-66). Urbana, IL: NCTE.

Mellon, J. C. (1969). Transformational sentence-combining: A method for developing the syntactic fluency in English composition. (NCTE Research Report No. 10). Urbana: IL: NCTE. Office of Management and Budget Circular A-110 to the Omnibus Supplemental Appropriations Act, Pub. L. No. 105-277. (1999).

O'Hare, F. (1973). Sentence-combining: Improving student writing without formal grammar instruction. (NCTE Research Report No. 15). Urbana: IL: NCTE.

Phelps, L. W. (1985). Dialects of coherence: Toward an integrative theory. *College English, 47,* 12-29.

Piaget, J. (1979). *Science of education and the psychology of the child.* New York: Penguin. Originally published 1969.

Purves, A. C. (1985). Apologia not accepted. *College Composition and Communication, 46*(4), 28-43.

Purves, A. C. (1988). Introduction. In A. C. Purves (Ed.), *Writing across languages: Issues in contrastive rhetoric* (pp. 9-21). Newbury Park, CA: Sage.

Rogan, P. (1999). Autonomy, interaction and emergence. Unpublished manuscript. Monterey Institute of International Studies.

Seltzer, R. (1998, October). Volunteerism and ESL make everyone a winner. Paper presented at the XV Rocky Mountain Regional TESOL Conference, Tucson, AZ.

Sheckley, B. G., & Keeton, M. T. (1997). Service learning: A theoretical model. In J. Shine. (Ed.) *Service learning: Ninety-sixth yearbook of the National Society for the Study of Education* (pp. 32-55). Chicago: National Society for the Study of Education.

Shohamy, E. (1998). Evaluation of learning outcomes in second language acquisition. In H. Byrnes (Ed.), *Learning foreign and second languages: Perspectives in research and scholarship* (pp. 238-261). NY: Modern Language Association.

Stotsky, S.L. (1975). Sentence-combining as a curricular activity: Its effect on written language development and reading comprehension. *Research in the Teaching of English, 9*, 30-71.

Thorndike, R. L., & Hagen, E. (1969). *Measurement and evaluation in psychology and education* (3rd ed.). NY: Wiley & Sons.

Toulmin, S.E. (1958). *The uses of argument.* Cambridge: Cambridge University Press.

Waldrop, M.M. (1992). *Complexity: The emerging science at the edge of order and chaos.* NY: Simon & Schuster.

Waller, T. (1993, March 31). Youth service can rebuild cities. *The Philadelphia Inquirer.*

White, E. M. (1995). An apologia for the timed impromptu essay test. *College Composition and Communication ,44* (2), 176-190.

White, E. M. (1998) *Teaching and assessing writing.* (2nd ed.). ME: Calendar Islands Publishers.

White, E. M., & Polin, L. (1986). Research in effective teaching of writing: NIE-G-81-0011 and NIE-G-82-0024. Washington, DC: National Institute for Education. (ED 275 007).

Wildner-Bassett, M. (n.d.). *Working Title: Beyond chaos: Research ideologies, intercultural pragmatics and positional pedagogies as complex dynamic systems.* Manuscript in preparation.

Witte, S. P. (1983a). Topical structure and revision: An exploratory study. *College Composition and Communication, 34*, 313-341.

Witte, S.P. (1983b). Topical structure and writing quality: Some possible text-based explanations of readers' judgments of students' writing. *Visible Language, 17*, 177-205.

Wurr, A. (1999). A pilot study of the impact of service-learning in college composition on native and non-native speakers of English. *Academic Exchange Quarterly 3(4)*, 54-61.

Wurr, A., Eröz, B., & Singh-Corcoran, N., (Eds.)(2000). *A Student's Guide to First-Year Composition.* (21st. ed.). Edina, MN: Burgess. In press.