

Adrian Wurr

## Text-Based Measures of Service-Learning Writing Quality

This paper describes methods to study the impact of service-learning on the writing performance of students in first-year college composition. Linguistic and rhetorical features commonly identified as affecting judgments of writing quality are compared to holistic essay ratings to assess the impact of different teaching and learning contexts on writing performance.

### Introduction

The rapid rise in the popularity of service-learning over the last decade has led to increased calls for evidence of its effect on student learning. Edward Zlotkowski, a senior associate at the American Association for Higher Education and author of numerous books and articles on service-learning, has repeatedly asserted that more academics would be swayed by empirical evidence showing gains in cognitive learning. Unless service-learning advocates become far more comfortable seeing enhanced learning as the horse pulling the cart of moral and civic values, and not vice versa, service-learning will continue to remain less visible and less important to the higher education community as a whole than is good for its own survival (Zlotkowski 24-25).

The study described here takes this charge seriously. As part of a larger dissertation study investigating the impact of service-learning on social, cognitive, and personal domains of learning, this paper describes a comprehensive writing assessment model used to measure the effects of service-learning on the writing performance of first-year college composition students. The model considers linguistic and rhetorical features in writing that can be compared to holistic evaluations of student writing and other qualitative assessments. Such direct measures of the writing produced by students in service-learning and comparison composition classes can help shed light on the potential academic benefits of service-learning in composition by answering the question: Does service-learning contribute to improved stu-

dent writing, and if so, in what ways?

#### Context

The service-learning curricula examined in the study were designed as part of a larger Southwest Project linking students and teachers at the University of Arizona with their counterparts in two local elementary schools to teach and learn about the land and people of the Southwest. Students in participating sections of first-year composition read and wrote about issues related to the Southwest in their college composition classes while also leading small group discussions on the Southwest in classes at two local elementary schools. The students were not only meeting community-defined needs by participating in this project, they were also meeting the goals of first-year composition by researching, designing, and drafting texts to meet the needs of multiple audiences.

The goals for first-year composition, as outlined in *A Student's Guide to First-Year Composition* and excerpted below, are as follows. Students will:

- Read texts to assess how writers achieve their purposes with their intended audiences.
- Learn the conventions of scholarly research, analysis, and documentation.
- Learn other conventions of academic writing, including how to write clear and correct prose.
- Learn to revise and respond to feedback from readers to improve and develop drafts
- Learn to develop ideas with observations and reflections on [their] experience.
- Learn to analyze and write for various rhetorical situations.
- Develop a persuasive argument and support it with evidence and effective appeals that target [their] intended audience. (Wurr, Er z, and Singh-Corcoran, 175)

Three major essay assignments—a rhetorical analysis, documented analysis, and reflective essay—are typically required of all first-year composition students and provide a basis from which to assess the degree to which curricular goals are met. In the service-learning courses, these assignments were designed to encourage students to reflect on the work in their community. Table 1 outlines the major writing assignments for service-learning courses in the study.<sup>1</sup>

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

- 
1. Rhetorical Analysis essay (5-7 pages), in which 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 introduces a portfolio of students accomplishments over the semester, and within which students explain why they chose the texts they did, for whom they are intended, and what purpose the texts or portfolio is meant to serve.
- 

Table 2 Language by Curricula Factorial Design

---

	NS (English 101)	NNS (English 107)
Service Learning	C1 (N=19)	C2 (N=16)
Comparison	C3(N=19)	C4 (N=19)

---

#### RESEARCH DESIGN

The primary participants in the study were a diverse mix of students enrolled in parallel first-year composition courses for native (NS) and non-native (NNS) English speaking students<sup>2</sup>. Students did not know about the service-learning component in these courses before enrolling, but were informed of this and other work related to the course in the first weeks of the semester. Each class section (labeled C1-C4 in Table 2) had a total enrollment of between 17 and 20 students<sup>3</sup> and one graduate student or adjunct faculty instructor.

Male and female participants were roughly equal in numbers and age, yet came from diverse linguistic and cultural backgrounds. Although a sample of convenience, the participants in this study are representative of the diverse student populations typically found on large, urban college campuses in America today.

Previous studies attempting to show gains in student writing as a result of teacher intervention and/or curricula have been largely unsuccessful in

part because much of what is taught in freshmen composition (e.g. research and library skills) is not easily documented or measured in the writing students produce (Haswell). A secondary purpose of the present study, therefore, was to investigate valid and reliable methods for describing writing quality based on current linguistic and rhetorical theories for analyzing student writing, with particular regard given to the persuasive essay writing emphasized in many writing courses incorporating service-learning. Holistic scores and primary trait analyses that had been used reliably to measure the use of rhetorical appeals, reasoning, coherence, and mechanics in other writing contexts were used as a means for documenting the impact of service-learning on student writing performance.

#### Holistic Writing Assessments

Holistic assessments of writing provide some advantages over primary trait scoring. In addition to accounting for the interaction of elements within a text, holistic or 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 than primary trait scales concerned with accuracy and mechanics to penalize writers, particularly second language writers, for surface level errors.

Since the essays in the present study were drafted over time, students had the opportunity to revise and edit all writing samples submitted for evaluation. This reduced the likelihood of fossilized errors appearing, as they often do in timed essay writing, by allowing students to avail themselves of a variety of resources including peer tutors, writing center consultants, the teacher, and computer grammar and spell check programs before they submitted their writing for evaluation. It is more likely, then, that the writing sample came closer to representing the student's true writing ability for the task.

All essays were rated by a team of qualified independent evaluators using the five-point scale presented in Table 3. Using exact and adjacent-scores, where up to a 1-point difference between scores is regarded as signifying agreement, the rubric resulted in an inter-rater reliability rate of .83 in the present study.

#### Analysis of Rhetorical Appeals

Ulla Connor and Janice Lauer 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,

Table 3 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.

the International Association for the Evaluation of Educational Achievement) conducted by Alan Purves 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 and Janice Lauer, Connor and Lauer describe measures for identifying and rating the use of three persuasive appeals: Rational, credibility, and affective, as shown in Table 4 ( Understanding ).

The IEA study achieved inter-rater reliability rates for the rational, credibility, and affective appeal scales of .90, .73, and .72 respectively (Connor, *Linguistic/Rhetorical* 76). As will be discussed more completely at the end of this paper, it is not clear from the literature if these figures represent exact

Table 4 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 picture, charged language, or metaphor to evoke emotion.
  - 3 Strong use of picture, charged language, or metaphor to evoke emotion.
- 

NOTE: From Connor, Ulla and Janice Lauer. "Cross-Cultural Variation in Persuasive Student Writing." *Writing Across Languages and Cultures*. Ed. Alan C. Purves. Newbury Park, CA: Sage, 1988. 138. Reprinted with permission.

or adjacent score agreement between raters. Using adjacent-score agreement, as is done in large-scale writing assessment programs such as the Test of Written English and others administered by the Educational Testing Service, the present study achieved inter-rater reliability rates of .94, .94, and .93 for the rational, credibility, and affective appeal scales respectively.

Analysis of Reasoning

In *The Uses of Argument*, Stephen Toulmin presents a model of informal logic to assess the soundness, strength, and conclusiveness of arguments that is comprised of three main parts: claims, data, and warrants (1). Claims are defined as conclusions whose merits we are seeking to establish (97). Data provide support for the claims in the form of experience, facts, statistics, or

**Table 5** 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 Connor, Ulla and Janice Lauer. "Cross-Cultural Variation in Persuasive Student Writing." *Writing Across Languages and Cultures*. Ed. Alan C. Purves. Newbury Park, CA: Sage, 1988. 138.

events. Warrants are rules, principles, [or] inference-licenses that act as bridges between claims and data (98). In *Cross-Cultural Variation in Persuasive Student Writing*, Connor and Lauer describe 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 5, Connor and Lauer's scale assesses both the quality and the quantity of the logic used. Using this scale, the present study achieved an average interrater reliability rate of .89.

## Analysis of Coherence

Research indicates that topical structure can be an important indicator of overall writing quality (Witte Connor, Contrastive; Connor and Farmer; Cerniglia, Medsker, and Connor). In *Topical Structure and Revision: Some Possible Text-Based Explanations of Readers Judgments of Students Writing*, Witte found that high quality essays had more parallel and extend-

Table 6 Bamberg's Four Point 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 Connor, Ulla and Janice Lauer. "Understanding Persuasive Essay Writing: Linguistic/Rhetorical Approach." *Text* 5.4 (1985): 311. ©Mouton Publishers. Reprinted with permission.

ed parallel progression than low quality essays.

Building on this idea, Betty Bamberg developed a system to help students revise their essays and improve coherence using topical structure



analysis. Connor and Lauer adapted this into a four-point rubric to measure text cohesion, and achieved an inter-rater reliability rate of .93 (Connor and Lauer 311). Shown in Table 6, Bamberg's system was chosen for the present study and resulted in an inter-rater reliability rate identical to that achieved by Connor and Lauer.

#### Analysis of Mechanics

Following Yili Li's dissertation study on the effect of computer-mediated communication activities on student writing, the present study measured the effective use of mechanics using the grammar checker of Microsoft Word 2000 (version 9.0.2720) to identify the following types of deviation from the conventions of standard academic English in the students' essays:

- Adverb: He writes bad.
- Article: A honest person would not do that.
- Capitalization: he took it.
- Comma use: It was late, the boys were hungry.
- Commonly confused words: Who's is that?
- Comparative: Writing is more easier than we expect.
- End-of-sentence preposition: She got the job she applied for.
- End-of-sentence punctuation: My bus had left.
- Extra word: The boat in the the basement was too big.
- Fragment: The rock samples.
- Negation: We couldn't hardly keep up with the orders.
- Number agreement: These banana are almost ripe.
- Parallelism: He should either pass or should bid.
- Possessives: The mans jacket was never found.
- Question mark: Who said that.
- Quotation marks: John said, I can't abide by that.
- Relative Pronoun: One person which I respect was Jim.
- Spelling (Confirmed manually): His assesment of the situation wasn't popular.
- Subject-Verb Agreement: One of the most important files are missing.
- Tense shift: He left and takes a nap.
- Verb Form: She had ran out of time.

The grammar checker on Microsoft Word 2000 was customized to identify only these types of errors. Because of the limitations of grammar checkers, manual confirmation was still necessary to avoid, for example, an in-text citation being counted as a fragment. The total number of confirmed errors in each student essay was divided by the total number of words to arrive at an

errors-per-word ratio. This ratio represented the level of adherence to formal conventions such as mechanics and style in a given essay. Using this method, the present study found the essays written in service-learning sections to have significantly fewer mechanical errors than those written in traditional comparison sections of first-year composition ( $p < .04$ ).

SUMMARY OF FINDINGS

The results from the holistic and analytic assessments of writing ability indi-

Table 7 Holistic and Analytic Writing Assessment Correlations (N=75)

	Holistic	Appeals	Logic	Coherence	Mechanics
Holistic Sig. (2-tailed)	1.00				
Appeals Sig. (2-tailed)	.337 .003	1.00			
Logic Sig. (2-tailed)	.507 .0001	.548 .0001	1.00		
Coherence Sig. (2-tailed)	.390 .001	.448 .0001	.509 .0001	1.00	
Mechanics Sig. (2-tailed)	-.219 .059	-.316 .006	-.338 .001	-.260 .024	1.00

cate a strong relationship exists ( $p < .01$ ) between a reader's holistic assessment of an essay's quality and the effective use of rhetorical appeals, logical reasoning, and cohesive devices in that same essay, as shown in Table 7. The high correlation between the primary trait and holistic scores suggests that rhetorical appeals, reasoning, coherence, and mechanics are not only significant variables to consider in assessing writing quality, but also important concepts to cover in teaching students how to become better writers. Composition students and teachers can benefit from gaining a more informed understanding of the most salient writing traits in holistic judgments of writing quality, while those more interested in service-learning can gain empirical support for their practices.

The study also found that the frequency of mechanical errors was inversely related to these same primary trait and holistic measures, suggesting the difficulty of assessing any one primary trait without some intervening

influence from other factors. Though the use of highly trained raters may have reduced this effect somewhat, the mild correlation between mechanics and other writing assessment measures raises the question of how best to account for such interaction when using analytic scoring mechanisms. In my dissertation, I suggest chaos theory can help account for such interactions through its description of complex and dynamic systems. I develop this idea by applying chaos theory to both writing assessment and service-learning program evaluation.

Table 8 Between group comparison of writing assessment scores

	Group	N	Mean*	Std. Deviation	Std. Error Mean
Holistic	SL	36	3.4722	.7923	.1321
	Control	39	3.0385	.8459	.1355
Appeals	SL	36	2.1750	.3544	5.906E-02
	Control	39	1.8810	.4214	6.748E-02
Logic	SL	36	2.2417	.3948	6.579E-02
	Control	39	1.7623	.3916	6.271E-02
Coherence	SL	36	3.1436	.5508	9.179E-02
	Control	39	2.6410	.6277	.1005
Mechanics	SL	36	3.16E-03	2.402524E-03	4.00E-04
	Control	39	4.67E-03	3.647087E-03	5.84E-04

\* NOTE: The range of possible scores for each variable above was 1-5 for holistic scores, 0-3 for appeals, 1-3 for logic, and 1-4 for coherence. Mechanics has an unlimited range from zero up, with means representing

Does Service-Learning Contribute to Improved Student Writing?

One of the most significant findings of the study was documenting a significant between-group difference on every writing assessment measure. The results shown in Table 8 indicate that independent raters judged the essays produced by students in service-learning sections of first-year composition to be superior to those produced in comparison sections in a variety of ways. These scores are mean scores for each group that were computed by averaging the scores individual raters assigned each essay on holistic and primary trait analyses. The scores given for rhetorical appeals and reasoning also represent mean scores for the three separate analyses and scores within each rubric (i.e., rational, credibility, affective rhetorical appeals, and claim, data, warrant scores respectively), but each of these components was

Table 9 Independent samples t-test for equality of means between groups

	T-Value	Degrees of Freedom	Significance (2-tailed)
Holistic	2.287	73	p<.025
Appeals	3.255	73	p<.002
Logic	5.276	73	p<.0001
Coherence	3.673	73	p<.001
Mechanics	-3.767	73	p<.04

scored separately by raters so that an essay could, for example, receive a score of 3 for warrants and a score of 1 for data.

Converting the 5-point holistic scale to letter grades, for example, reveals that service-learning essays were judged to be better than comparison essays by about half a letter grade. Analytic assessments of each group's use of rhetorical appeals, logic, coherence, and mechanics show service-learning essays to be superior to comparison essays on every measure. The data presented in Table 9 indicate that the possibility of these results occurring by chance was small.

These results provide empirical support for including service-learning in college composition curricula. While other studies have demonstrated the positive impact service-learning can have on the community (e.g., Gelmon et al. and Gray et al.), this study has shown that incorporating service-learning in college composition improves student writing, and provided a viable model for assessing this growth in student writing.

#### Suggestions for Future Research

The primary trait and holistic assessments of the writing produced by students in service-learning and comparison sections of first-year composition were conducted with readers familiar with the institutional context of the present study and with the institution's writing assessment procedures. As a result, inter-rater reliability rates were .83, .94, .89, and .93 for holistic, rhetorical appeals, reasoning, and coherence respectively. These results compare favorably to those reported in other studies (see Table 10) that involved larger samples drawn from diverse instructional and cultural settings. The range of writers and writing contexts represented in Connor and Lauer's two studies, for example, may have complicated the assessment task for their raters. The cohesiveness of the educational setting in which the present study occurred may have had a positive impact on the outcomes and

Table 10 Inter-rater reliability rates from the present and comparison studies

	Exact match	1-point difference	Comparison study
Holistic	.38	.83	.83b
Appeals	.50	.94	.78b
Logic	.54	.97	.66c
Coherence	.47	.93	.93a

NOTE: a = Connor and Farmer; b = Connor and Lauer "Understanding"; c = Connor and Lauer "Cross-Cultural"

help account for the high interrater-reliability rates achieved (White).

It is important to note, however, that these inter-rater reliability rates are for adjacent score matches. When inter-rater scores were two or more points apart, the raters were said to have disagreed in their assessment of the essay and a third reader would read the essay. An average of all three raters scores would comprise the essay's final assessment score. Since it is not clear from other published studies using similar holistic and primary trait instruments exactly how they determined their reported inter-rater reliability rates, I have included both sets of figures in Table 10. The results from the comparison studies mentioned above are listed in the third column. Inter-rater reliability rates for mechanics are not included since these were calculated by computer.

Each rating session in the present study began with an explanation of the assessment procedure and scoring guide to be used, followed by practice in scoring several sample essays in order to reinforce the points outlined on the scoring guide. Nevertheless, some problems arose in interpreting and applying at least two of the rubrics. The 3-point scales for judging the quality of claims, data, and warrants in the analysis of reasoning rubric all seemed to contain a gap between scores 1 and 2. One rater, for example, asked how she should score an essay with an easily identifiable argument structure, but one in which the writer's ethos undermined the argument's credibility. After discussing the point, all three raters seemed to agree in word but not deed as their differing scores on such essays led to several 1-3 splits, particularly on claims and warrants. The 4-point rubric for analyzing coherence was also problematic. Only one of the 75 papers in the study received a score of 1 since the descriptor, essay is literally incomprehensible, rarely applies to college level writing; softening this criterion to mostly incomprehensible

might be more effective and appropriate. Also, with the practical elimination of a score of 1 from the coherence scale, many essays ended up receiving a score of 3 since this represented the middle ground between two more thoroughly defined alternatives. The descriptors for scores of 2 and 4 could benefit from being defined more narrowly, and the descriptor for 3 needs to be described in more detail, not just in opposition to adjacent scores. Although refining the rubrics in these or other ways more applicable to the local context may lead to higher inter-rater reliability rates, some variance in scoring is inevitable given the individual nature of reader responses.

#### Process vs. Product

One drawback of focusing on text-based measures of writing quality is that it may have the undesirable effect of emphasizing written products over composing processes. It would be interesting to investigate the composition strategies students in campus and community-based writing courses used as they completed their writing assignments. Protocol analysis might be one way in which to investigate this topic (see, for example, Flower and Hayes; Penrose and Sitko).

Another limitation of the writing assessment model described here is that focusing on a single or even multiple writing sample collected within a semester may not capture long-term effects on writing performance. One would hope to see changes in participants attitude and behavior concerning writing as well as in the actual writing produced before concluding that a particular course or methodology had an impact on student writing performance. To address this concern, teacher-researchers might also consider the students own self-assessment of their writing development as a result of the course, as indicated by reflective journal writing, writing portfolio, end-of-the-semester course evaluations, and interview data. All of these components were included in my dissertation study, but space limitations prevent describing them in detail here. The results, however, were generally encouraging with regard to the positive impact service-learning can have on students beliefs about writing and research. A longitudinal, multi-institutional study of service-learning might reveal the extent to which these results accurately describe the outcomes for writers working in other service-learning teaching and learning contexts

#### Notes

<sup>1</sup> The writing portfolio mentioned in assignment #3 consisted of journal assignments, research and writing done for community partners, and a multi-draft-

ed essay. These portfolios were scored holistically using a 6-point rubric. The holistic scores for writing portfolios and persuasive essays were strongly correlated, so only the persuasive essay scores and rubric are discussed in this paper. See my dissertation for more on the writing portfolios (133-4).

<sup>2</sup>See Wurr ( We are ) for more on the comparison of the impact of service-learning on NS versus NNS students

<sup>3</sup>One student in C2 and C4 declined to participate in the study, hence the difference in the number of enrolled students and study participants

#### Works Cited

- Bamberg, Betty. What Makes a Text Coherent? *College Composition and Communication* 34 (1983): 417-29.
- Cerniglia, C., K. Medsker, and Ulla Connor. Improving Coherence Using Computer-Assisted Instruction. *Coherence in Writing: Research and Pedagogical Perspectives*. Eds. Ulla Connor and A. M. Johns. Arlington: TESOL, 1990. 227-41.
- Connor, Ulla. Linguistic/Rhetorical Measures for International Persuasive Student Writing. *Research in the Teaching of English* 24 (1990): 67-87.
- . *Contrastive Rhetoric: Cross-Cultural Aspects of Second Language Writing*. NY: Cambridge UP, 1996.
- , and Mary Farmer. The Teaching of Topical Structure Analysis as a Revision Strategy for ESL Writers. *Second Language Writing: Research Insights for the Classroom*. Ed. Barbara Kroll. NY: Cambridge UP, 1990. 126-39.
- , and Janice Lauer. Understanding Persuasive Essay Writing: Linguistic/Rhetorical Approach. *Text* 5.4 (1985): 309-26.
- , and . Cross-Cultural Variation in Persuasive Student Writing. *Writing Across Languages and Cultures*. Ed. Alan C. Purves. Newbury Park: Sage Publications, 1988. 138-59.
- Flower, Linda, and John R. Hayes. Images, Plans, and Prose: *The Representation of Meaning in Writing*. *Written Communication*, 1.1 (1984): 120-160.
- Gelmon, Sherril B., et al. (1998). Community-University Partnerships for Mutual Learning. *Michigan Journal of Community Service Learning* 5 (Fall 1998): 97-107.

Gray, Maryann J. Assessing Service-Learning: Results from a Survey of Learn and Serve America, Higher Education. *Change* (March/April 2000): 30-39.

Haswell, Richard. *Gaining Ground in College Writing: Tales of Development and Interpretation*. Dallas: Southern Methodist UP, 1991.

Kinneavy, James L. *A Theory of Discourse*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

Lauer, Janice M. et al. *Four Worlds of Writing*. 2nd ed. New York: Harper and Row, 1985.

Li, Yili. Using Task-based E-mail Activities in Developing Academic Writing Skills in English as a Second Language. Diss. U of Arizona, 1998.

Penrose, Ann M and Barbara M. Sitko, eds. *Hearing Ourselves Think: Cognitive Research in the College Writing Classroom*. Oxford: Oxford UP, 1993.

Purves, Alan C. *Writing Across Languages and Cultures: Issues in Contrastive Rhetoric*. Newbury Park, CA: Sage, 1988.

Toulmin, Stephen E. *The Uses of Argument*. Cambridge: Cambridge UP, 1958.

White, Edward M. *Personal Interview*. 11 Nov. 2000.

White, Stephen P. Topical Structure and Revision: An Exploratory Study. *College Composition and Communication* 34 (1983): 313-41.

Wurr, Adrian J. The Impact and Effect of Service-Learning on Native and Non-Native English Speaking College Composition Students. Diss. U of Arizona, 2001.

Wurr, Adrian J. We Are (Not) One: Managing Diversity in Community-Based Writing Programs. Paper to be Presented at the Annual Conference on College Composition and Communication, Chicago, IL 2002.

Wurr, Adrian J., Betsy Erz, and Nathalie Singh-Corcoran, eds. *A Student's Guide to First-Year Composition*. 21st. ed. Edina, MN: Burgess, 2000.

Zlotkowski, Edward. Linking Service-Learning and the Academy: A New Voice at the Table? *Change* (January/February 1996): 21-27.

Adrian Wurr teaches courses in applied linguistics and composition at UNC Greensboro. His research interests include TESOL, service-learning, and literacy instruction. He is currently collaborating on an ESL Composition book that will be published by Lawrence Erlbaum Associates.



Automated text assessment systems exploit textual features chosen in an attempt to balance evidence of writing competence against evidence of performance errors in order to measure the overall quality and assign a score to a text. The earliest systems used superficial features, such as word and sentence length, as proxies for understanding the text. More recent systems have used more sophisticated automated text processing techniques to measure grammaticality, textual coherence, prespecified errors, and so forth. In the next section, we provide an overview of this thesis, followed by our research