

Lessons about Writing to Learn from a University-High School Partnership

Bradley Peters

ABSTRACT

This essay focuses on a university-high school partnership that emphasized the design and implementation of writing-to-learn tasks in the curriculum of a low-income school on “Academic Watch.” A university coordinator of writing across the curriculum (WAC) teamed up with a special education high school instructor to teach a professional development course, which was sponsored by a federal Teacher Quality Enhancement grant. Replicating National Writing Project research methods, a follow-up exploratory study yielded statistically significant quantitative data and supportive qualitative data that together demonstrated how writing to learn exerts positive effects on student learning outcomes (“NWP Research Brief”). In its emphasis on quantitative data, the study sought to avoid what meta-analysts have identified as problems of reliability, validity, and control in research on writing. The study’s findings contribute to what such analysts call a more “pedagogically useful theory” for implementing writing to learn in secondary and post-secondary settings.

I. PROBLEMS WITH STUDIES OF WRITING TO LEARN

The history of scholarship in WAC espouses informal, write-to-learn tasks as the supplementary or default method for integrating writing in cross-disciplinary courses—especially when faculty in universities and high schools shy away from assigning written projects that require research, multiple drafting, and feedback (e.g., Fulwiler 24-25; Zinsler 154-56; McLeod 4; Walvoord, et al. 91-92, 100-102; Bean 97-98; Duffy 118-121). Recent work by the Consortium for the Study of Writing in College (CSWC)—a joint project between the Council of Writing Program Administrators (CWPA) and the National Survey of Student Engagement (NSSE)—reiterates the

importance of writing to learn in a new light. In 2008, researchers administered 27 supplemental questions about writing practices in the NSSE. They drew response from approximately 23,000 students in 82 four-year, post-secondary institutions selected at random. As the researchers assert:

Results suggest that faculty can increase student engagement in deep learning activities and also increase student learning by including interactive activities, assigning meaning-constructing writing projects, and clearly explaining their expectations. The results also suggest that these factors contribute more to the achievement of desirable learning outcomes than does the amount of writing faculty assign (Anderson, Anson, Gonyea, and Paine).

Based on the CSWC's indirect measures, the implications for writing to learn could persuade many more cross-disciplinary faculties to incorporate writing into their pedagogies for the benefit of their students.

However, two contemporary meta-analyses of research on writing to learn warn against oversimplifying such claims. The authors of these meta-analyses, hailing from educational theory, practice, and curriculum development, find that the act of writing itself "does not automatically yield large dividends in learning" (Klein 206; Bangert-Drowns, et al. 29). Both meta-analyses suggest that to get the promising results that the CSCW recounts, proponents must heed four theoretical claims which have shaped our current understanding of writing to learn (Klein 211; Bangert-Drowns, et al. 30, 32-33).

First, to deal with unfamiliar materials, students need to use writing to generate response, personally translate, and collectively share their thoughts. Such writing, "being the form of writing nearest speech, is crucial for trying out and coming to terms with new ideas" (Martin, et al. 43; also see Britton; Vygotsky). Second, to acquire and practice the critical thinking that structures relationships among ideas, students should develop a repertoire of learning strategies such as defining, classifying, explaining cause-effect, and substantiating claims (e.g., Emig; Bereiter and Scardamalia; Langer and Applebee; Newell; Newkirk). Third, to transform unfamiliar material into knowledge, students must review their initial writing, so as to elaborate upon, evaluate, organize, and revise what they've produced (e.g., Bruner, Bereiter and Scardamalia; Flower and Hayes, 1980, 1981). Fourth, to develop the metacognitive skills characteristic of self-directed learners, they must write to reflect upon "their current understandings, confusions, and feelings in relation to the subject matter" (e.g., Bangert-Drowns, et al. 47; McCrindle and Christensen; Yancey). Both meta-analyses agree that to formulate "a more pedagogically useful theory of writing to learn," class-

room practices must draw upon each of these four theoretical claims in complementary fashion (Klein 255; Bangert-Drowns, et al. 50).

At the same time, both meta-analysts note that most empirical studies of writing to learn do not elaborate sufficiently on how instructors are trained to apply these four claims, how students are taught to compose with them, or how the students' social context affects their composing (Klein 207; Bagert-Drowns, et al. 36). Furthermore, they note that even the best studies of writing to learn lack control groups, pre/post measures of learning, consistent methods, or the detail that compelling ethnographic research requires (Klein 205; Bangert-Drowns, et al. 39, 40-41). According to this criticism, a dearth of quantitative evidence confirming benefits of writing to learn keeps WAC proponents from validating a rigorous theory that manages to "generalize across task content," posit strategies that "fall within the zone of proximal development of many students," and "include readily teachable strategies" (Klein 255).

Within composition studies, disciplinary critics agree. Portfolio experts Liz Hamp-Lyons and Bill Condon observe that even the assessment of a more robust model of writing—including multiple drafts, peer review, instructor feedback, revision, and editing—relies upon "the traditional crude measures of interrater reliability and criterion validity" (136-137). Tougher questions about reliability remain unanswered. Richard Haswell, co-author of *CompPile*, the discipline's most comprehensive publication data base, reports a paucity of "replicable, aggregative, and data-supported" studies on theories of writing pedagogy (2005, 210). He notes that neither the National Council of Teachers of English nor the Conference on College Composition and Communication encourages such studies. Doug Hesse, former program chair of CCCC, and Chris Anson, past president of the CWPA, both warn that "the lack of compelling findings on pedagogies" undermines the power of writing instructors and administrators "to respond critically to reports of research that will be used to decide how they will teach, what they will teach, and to what ends" (Hesse W421; Anson 28).

The following pages detail an exploratory project that begins to address these problems. I start with brief descriptions of how the project was supported, where the project took place, who the participants were, and how the professional-development course was designed. An account of the follow-up assessment ensues, in terms of who took part in it, how subjects were selected, what procedures were established, and how the teachers were prepared for the final rating of student learning outcomes. An explanation of what quantitative data were collected and how those data were analyzed comes next—substantiating why such data might matter to high school

teachers, composition instructors, cross-disciplinary faculty, writing program administrators, and other stakeholders. This essay ends with a discussion of how certain qualitative data help illuminate the quantitative data further, indicating where the follow-up study's more promising results—as well as its limitations—open the way for improved research design, methods, and future meaningful investigation.

II. A WAC COURSE IN A ROUGH ENVIRONMENT

This project began when my university's College of Education responded to the federal call of No Child Left Behind to intensify coordination between higher education and school districts. The Dean successfully applied for a five-million dollar Teacher Quality Enhancement (TQE) grant sponsored by the US Department of Education. She invited the College of Liberal Arts and Sciences—where secondary teacher certification programs are housed—to participate. The TQE grant would fund a partnership with an urban district, where several of the professional-development schools at the elementary and secondary levels were on Academic Watch. As WAC coordinator and instructor of a teacher certification course in writing methods for middle and high school, I saw a promising opportunity to implement National Writing Project methods and materials without also getting caught up in the extensive fundraising, cost-sharing, and accounting practices that federal law requires (NCLB, Title II, Part C, Ch. C, Subpart 2, Sect. 2332).

Setting

The professional-development high school that the TQE grant identified had a poverty-income index of 58%, with an ethnic mix of 28% black, 20% Hispanic, 47% white, and 5% “other” (“Great Schools”). The potential for ethnic tensions—in combination with egregious inequities in state school funding¹—contributed “to the social turbulence of adolescents in so many schools like this” (Kozol 28). About 25-30% of the teachers transferred out annually (*JHS Improvement Plan 3*). One semester during the school's involvement with the TQE grant, nearly three-quarters of the math department resigned. Another semester, a reading teacher was thrown against a wall when she attempted to stop a fist-fight. After she got out of hospital, she never returned.

Administrators used the same revolving door. Four different principals served throughout the grant's five years. The district transferred one principal to another school. In a highly racialized controversy, another was fired because he changed 842 failing grades to passing, to improve the school's

annual state report card (“Preliminary Report” 9). Two bomb threats, student protests, and school closings occurred while a substitute principal tried to restore order. Peace only came when the fourth took over.

Participants

Experienced participants in university-high school collaborations elsewhere assert that teachers need the opportunity “to experience education as a working continuum, not as a fragmented system in which their individual voices cannot be heard” (Morris, et al. 169). Participants in this project were all too familiar with such fragmentation. Because this professional-development school was on Academic Watch, district administrators required teachers to take several weeks away from content instruction every semester, to have students practice as well as take the requisite battery of standardized tests. The tests often did not coincide with what teachers taught, e.g., students had test questions on earth science, but no high school courses in earth science (*Prairie State Achievement Examination 55-59*; “Science Curriculum”). To teachers’ further frustration, district administrators micromanaged curriculum, imposing new test-score improvement programs upon all content areas, sometimes on a per-semester basis.

Surprisingly, when several teachers in the high school met to discuss implementing WAC as part of the TQE grant’s objectives, they showed keen interest. In particular, they speculated that integrating writing-to-learn activities might provide a consistent method to shore up the teaching of content in their subject areas, offsetting—as one teacher put it—the school district’s predilection for “prescribing fad-of-the-month gimmicks to bolster test scores.” We agreed upon three main goals:

- Offer an on-site course to teach writing in various subject areas
- Work with faculty to enhance writing/ reading in their individual classes
- Coordinate with faculty to design authentic assessment of writing

A special education reading teacher with strong experience, energy, and commitment to writing joined me as course co-facilitator. During the next three years—despite the dramatic distractions—31 teachers from 11 different disciplinary areas enrolled (zoology, business, economics, English, French, history, physical education, math, music, Spanish, reading, and special education). Three administrators participated as well: a counselor, a truant officer, and one of the principals. This enrollment represented 20% participation from a high school with 100 full-time teachers and 21 para-

professionals. The course turned out to be the single-most sustained part of the TQE grant's high school involvement, while unrepeated classes or single workshops predominated (*JHS Improvement Plan* 18-21).

Course Design

We based the WAC course on four best practices identified by the 1998 National Assessment of Educational Progress. The practices corresponded closely to the four theories that this essay mentioned previously:

- Get students to plan informal writing at least twice a month
- Discuss students' writing strategies with them, in the context of course content
- Have students do some form of redrafting
- Ask students to collect their writing, so they can examine and reflect upon what they've learned (Nagin 44)

The teachers themselves engaged in these four practices. They welcomed the notion that assigning long, formal papers was unnecessary (see Bean 117). Course activities involved developing a series of four sets of closely linked, informal prompts for a semester. Each set asked students to write three or four brief, impromptu responses about a central concept that students needed to learn. Each impromptu response comprised a separate day's 5-10 minute writing activity. An interactive discussion or reading assignment about the concept preceded or followed. Each prompt encouraged students to engage in what teachers defined as one or two specific learning strategies, leading up to a slightly longer 10-15 minute activity that required students to redraft and synthesize what they'd written before. A short, reflective prompt wrapped up each set. Just so, this sequenced approach applied the four theoretical claims of writing to learn.

To illustrate, a zoology teacher developed this set of prompts on flatworms:

- Write a paragraph about general traits of the phylum Platyhelminthes, and describe traits that differentiate the four classes.
- Choose two parasitic flatworms that we've read about or studied online, comparing their life cycles. Identify intermediate and definite hosts, and the stage of development that the flatworms are found in each host. Then draw two comparative charts, using scientific and common names.

- Choose a third parasitic flatworm that we've researched. Write about the disease that it causes, where and how people 'catch' the disease, what symptoms occur, how the disease is cured, and how you'd avoid it.
- With two other students, write a public health pamphlet informing a community about the health problems that one specific species of flatworm cause. Provide information about its life cycle, prevention, symptoms, and public health sources. Include photos and drawings.
- What were the three most important things you learned from writing about flatworms? Identify what was most difficult, and explain why. What did we do that was most helpful to you? Again, explain why.

Although we didn't suggest that teachers grade student responses to these prompts, we developed a rubric that incorporated state standards, to evaluate if students met expectations, missed, or excelled (see Elbow 86-87). The rubric's third criterion also included learning strategies that Klein and Bangert-Drowns, et al. had identified.

Table 1: Rubric (Score = sum of ratings for each criterion divided by 4 and rounded to nearest tenth)

CRITERIA	3-exceeds expectations	2-meets expectations	1-misses expectations
Comprehension of Task —ability to respond informally to what a writing prompt asks	Student always understands and follows instructions exactly.	Student usually understands and follows instructions.	Student often misunderstands or disregards instructions.
Content —ability to convey knowledge of course content obtained from reading or listening	Student provides information that is accurate and detail that supports it very well.	Student provides information that is mostly accurate and detail that is adequate.	Student provides information that is not accurate and/or detail that is insufficient.
Strategies —ability to apply, analyze, back up, compare, classify, critique, define, describe, evaluate, explain, exemplify, illustrate, interpret, question, reflect, review, show cause-effect, solve, synthesize, translate, etc.	Student shows clear control over the strategy or strategies that the prompt requires.	Student shows satisfactory evidence of understanding and practicing the strategy or strategies required.	Student shows little or no evidence of understanding the strategy or strategies required.

CRITERIA	3-exceeds expectations	2-meets expectations	1-misses expectations
Language usage —ability to develop a readable response and to use conventions of grammar and punctuation	Student’s response is very articulate, and errors are too minimal to worry about.	Student’s response is fairly easy to read, and errors do not keep me from understanding.	Student’s response is illegible and/or difficult to understand, and errors confuse me.

At the end of the course, the teachers reviewed their folders and reflected on what was especially valuable about designing and incorporating these informal, write-to-learn tasks into their curricula. The zoology teacher said: “The most valuable part of this second-semester project was putting together more than one linked set of writing assignments... I have learned a very different approach to getting students to write.” A math teacher noted that he now wanted to use writing to help students realize how course content was applicable to their lives—and “not just another hoop to jump through for that doggy treat of a diploma.... To make this work, I have to rethink how and what I am teaching.” An English teacher offered the following comments:

Writing good prompts depends on many different factors—breaking a task down, audience, timing, talking about yourself as a writer, helping students develop a knowledge base, explaining terms, and a classroom’s behavioral context. I know a lot of these ideas already, but making me think about them, sort them out, and use them has been really beneficial.

III. THE FOLLOW-UP ASSESSMENT

Although the WAC course had encouraged teachers to rethink their teaching practices, my co-facilitator and I wondered who would go on applying the principles they’d learned, and what outcomes might emerge. Four of our alumni from zoology, economics, English, and history volunteered to give at least a year to do an assessment project. The co-facilitator added reading to the mix. Later, the history teacher had to withdraw for personal reasons.

Selection of Subjects and Procedures

The follow-up project drew students from various disciplines and grade levels (9 through 12) who engaged in sustained, impromptu writing and

assembled it in folders that we assessed. Students did a final reflection letter that:

- Explained which they thought was their strongest example and why
- Critiqued which was weakest and why
- Defined which three concepts they felt they had learned best
- Reflected on how writing had helped them learn those concepts

Participating teachers identified one class where students would do sustained, informal writing and another, like class where students would not. Students in both the “writing-folder” and “non-folder” classes completed all other coursework (e.g., worksheets, lab reports, term papers, collaborative projects, exams). They all wrote a brief piece before and after their coursework, so we could gauge if the writing-to-learn groups had made any gains. For instance, “pre-course” economics students in both the writing-to-learn and control classes wrote about a personal problem (financial or otherwise) they had worked through. For the “post-course” activity, the same students identified what they thought was America’s most compelling economic problem, how it affected them or their families, and how the country might work through it to achieve a stronger recovery from the recession. Such pre-/post- writing samples are valid and may show statistically significant gain “if exact measures match what teachers are actually teaching and students actually practicing in their writing” (Haswell, “Assessment”).

The teachers from senior economics and junior zoology participated for one year, while the teachers from sophomore AP English and freshmen special education reading volunteered for two. During the first year, I observed classes in zoology, English, and reading, to get a clear idea of classroom dynamics. I also visited an economics class once. Both years, we all met every two weeks to compare several factors, including:

- Design of the sequenced writing prompts
- Feedback students received (commentary and primary-trait scoring²)
- Gains that students ostensibly made from writing to learn
- Progress in the “folder” and “non-folder” classes

All teachers wrote quarterly reflections to detail what insights they were gaining about their students’ learning outcomes and their own teaching. For one semester, a student teacher in economics joined our bi-weekly discussions, so he could understand the project and consistently administer the prompts that the economics teacher designed.

Preparation and Final Assessment

Every quarter during the two years, we rated sample folders from a stronger and weaker student in each course. The teachers and I used the rubric shown in Table 1, based on the three-point scale. Then we discussed how and why we rated the four criteria.

At the end of each year, two teachers other than the teacher of record rated the teacher of record's folders (e.g., the special education and English teachers rated zoology folders). For the first year, two teacher-certification specialists from my university—one in history and one in English—also rated each class set. I served as a rater for both years. Being the only reader who had rated all student work, I also used the rubric to rate the pre/post writing samples from all folder and non-folder classes. After rating the pre/post writing, I checked my scores against the pre/post scores that each teacher of record had assigned. Finally, each teacher wrote a reflection about what they had learned that year.

Under these protocols, we hoped to control not only for validity, as researchers advise, but also for reliability, given the primary variables of grade level, subject area, length of participation, and ability grouping (see Hamp-Lyons and Condon 137-138).

IV. FINDINGS

The quantitative data in this section provide insights about student learning outcomes in ways that indirect methods or qualitative data cannot. Brief rationales accompany each statistical measure, to validate assessment methods and to explain the extent to which writing to learn may have exerted a positive effect.³

Basic Data and Statistical Measures

Raters assessed a total of 88 folders: 49 at the end of the project's first year and 39 at the end of the second. In 2008, the three high school teachers, two university teacher certification specialists, and I rated each of 49 folders for a total of 245 scores. In 2009, the three high school teachers and I rated each of 39 folders for a total of 117 scores. In addition, I rated 61 pre-/post-course writing samples from the folder group and 55 pre/post samples from the non-folder group.

Pearson r was calculated to correlate interrater reliability between different pairs of raters. Means and standard deviations were calculated for scores on each of the four rubric criteria—comprehension of task, content, strategies, and language usage—to find the students' collective central

tendencies as well as their central tendencies in each subject area. Percentages were calculated for scores on each of the rubric's criteria to compare the students' overall dispersion of written performances and the dispersion of their written performances in each different subject area, grade level, and ability group. A *t* test for unequal samples was calculated between the scores for the pre-/post-course writing samples, both for the folder groups and the non-folder groups, to see: (A) what differences existed between the groups, and (B) if those differences might be statistically significant (see Steinberg 235). Effect size was calculated to judge if any statistical significance between scores for pre-/ post-course writing samples might be large enough to claim that writing to learn indeed did make an impact on the folder groups' learning outcomes. Finally, means, standard deviations, and percentages were calculated for each rubric criterion in the pre-/post-course writing samples, to compare in more detail where the folder and non-folder groups' strengths and weaknesses lay.

Interrater reliability

Pearson *r* is a common calculation to determine correlations between pairs of scores.⁴ Tables 2 and 3, below, identify the Pearson *r* correlations for each pair of raters. To illustrate, the Table 2 correlation of scores between raters NC and LG in special education reading 2008, which had 9 students, is $r(7) = .88, p < .01$. This result shows that Pearson *r* at 7 degrees of freedom is $+0.88$, with less than a 1% probability, *p*, that the correlation of scores is due to chance (see Steinberg 409). To state that the raters agreed upon the benefits that students in special education reading gained from writing to learn is overly optimistic, however, because 9 constitute such a small sample size.⁵ A more convincing correlation exists between DS and BP (me) in zoology 2008, where $r(18) = .91, p < .01$. The result comes from a larger group of 20 students (18 *df*), showing that there is less than a 1% probability, *p*, that the correlation occurred by chance.

Interrater reliability was strongest and most consistent for the zoology 2008 and economics 2009 folders. It was the weakest and least consistent for English 2008 and special education reading 2009. The seemingly high 2009 special education correlation coefficients yield lower or no statistical significance because there were only 6 students who completed folders by the end of that year. As seen, a fairly diverse range of correlations showed up among pairs of raters. Yet many more correlations were statistically significant than not, with *p* ranging from $<.01$ to $<.05$. I tended to give the lowest ratings, accounting for three of the Pearson *r* coefficients that failed to demonstrate statistical significance. The university teacher certification

Table 2. Pearson *r* Calculated for Pairs of Readers, Year One

SPED READING 2008			ZOOLOGY 2008			ENGLISH 2008		
Raters	Correlation of Scores		Raters	Correlation of Scores		Raters	Correlation of Scores	
NC/LG	$r(7) = .88, p < .01$		DS/BP	$r(18) = .91, p < .01$		JP/JH	$r(18) = .82, p < .01$	
LG/JH	$r(7) = .86, p < .01$		DS/LG	$r(18) = .87, p < .01$		DS/JP	$r(18) = .67, p < .01$	
JP/JH	$r(7) = .74, p < .05$		DS/JH	$r(18) = .78, p < .01$		DS/JH	$r(18) = .66, p < .01$	
BP/JH	$r(7) = .72, p < .05$		DS/JP	$r(18) = .77, p < .01$		DS/BP	$r(18) = .51, p < .05$	
LG/BP	$r(7) = .70, p < .05$		LG/BP	$r(18) = .77, p < .01$		BP/JP	$r(18) = .49, p < .05$	
LG/JP	$r(7) = .69, p < .05$		LG/JH	$r(18) = .74, p < .01$		NC/DS	$r(18) = .42, \text{not significant}$	
NC/BP	$r(7) = .68, p < .05$		JP/JH	$r(18) = .72, p < .01$		NC/JH	$r(18) = .41, \text{not significant}$	
BP/JP	$r(7) = .67, p < .05$		BP/JH	$r(18) = .69, p < .01$		NC/JP	$r(18) = .39, \text{not significant}$	
NC/JH	$r(7) = .67, p < .05$		LG/JP	$r(18) = .69, p < .01$		BP/JH	$r(18) = .37, \text{not significant}$	
NC/JP	$r(7) = .48, \text{not significant}$		BP/JH	$r(18) = .66, p < .01$		NC/BP	$r(18) = .32, \text{not significant}$	

Table 3. Pearson r Calculated for Pairs of Readers, Year Two

SP ED READING 2009		ECONOMICS 2009		ENGLISH 2009	
Raters	Correlation of Scores	Raters	Correlation of Scores	Raters	Correlation of Scores
DC/LG	$r(4) = .91, p < .02$	DS/LG	$r(15) = .88, p < .01$	DC/DS	$R(14) = .85, p < .01$
DC/BP	$r(4) = .87, p < .05$	DS/BP	$r(15) = .64, p < .01$	DC/BP	$R(14) = .61, p < .01$
LG/BP	$r(4) = .72, \text{not significant}$	LG/BP	$r(15) = .58, p < .01$	DS/BP	$R(14) = .59, p < .02$

specialists did not join the second year of assessments because funds were cut back. Further comments on these and other results appear in the discussion section that follows.

Central Tendencies and Dispersion of Scores

Calculating central tendencies and the dispersion of scores makes it easier to aggregate and compare groups of different sizes, by “mapping” results back to criteria such as those in the rubric used for this project (Suskie 258-260).

Table 4, below, provides means, standard deviations, and a percentage dispersion of scores. An immediate good sign that writing to learn might have had a positive impact is that the means for the rubric’s individual criteria, as well as overall scores, tended to range no more than .5 beyond 2, the point for meeting expectations. The means tended to be lowest and standard deviations highest with the most challenging criteria—content accuracy and learning strategies. Comprehension of written tasks generally appeared to present less of a challenge, as did language usage. In general, the means for AP English tended to skew higher as might be anticipated.

An aside: it is preferable to calculate standard deviation with no less than 30 subjects. But in these data sets, calculating standard deviation for the means of smaller groups seemed appropriate and instructive because the numbers satisfied another important condition: overall dispersion of percentages in each subject area, as well as in all subject areas combined, approximated a typical bell curve (see Steinberg 453). Standard deviation was also necessary for calculating the overall percentage dispersions.

Table 4. Scores analyzed by Means, Standard Deviations, and Percentages

Student Learning Outcomes	Mean	S.D.	Exceeds	Meets	Misses
All subjects (Grades 9-12, two years, n=88, 362 readings)					
Comprehension of task	2.17	.58	27%	63%	10%
Content	1.98	.69	23%	52%	25%
Strategies	2.09	.69	29%	51%	20%
Language usage	2.25	.66	38%	50%	12%
OVERALL	2.15	.54	23%	58%	19%
AP English (Grade 10, two years, n=36, 148 readings)					
Comprehension of task	2.43	.51	43%	56%	1%
Content	2.36	.56	40%	56%	4%
Strategies	2.5	.55	52%	45%	3%

Student Learning Outcomes	Mean	S.D.	Exceeds	Meets	Misses
Language usage	2.55	.49	55%	45%	0%
OVERALL	2.48	.40	20%	73%	7%
Zoology (Grade 11, one year, n=20, 100 readings)					
Comprehension of task	1.95	.56	13%	69%	18%
Content	1.64	.67	11%	42%	47%
Strategies	1.77	.69	15%	47%	38%
Language usage	1.99	.64	19%	60%	21%
OVERALL	1.86	.52	17%	63%	20%
Economics (Grade 12, one year, n=17, 51 readings)					
Comprehension of task	2.18	.51	23%	71%	6%
Content	1.98	.61	18%	63%	19%
Strategies	1.94	.65	17%	59%	24%
Language usage	2.62	.49	63%	37%	0%
OVERALL	2.21	.45	19%	59%	22%
Sp. Ed. Reading (Grade 9, two years, n=15, 63 readings)					
Comprehension of task	1.94	.59	14%	65%	21%
Content	1.66	.62	8%	51%	41%
Strategies	1.75	.54	5%	65%	30%
Language usage	1.65	.51	1%	62%	37%
OVERALL	1.77	.43	14%	67%	19%

Setting up a percentage dispersion of scores for each criterion provides a more detailed comparison of how the students performed in their different subject areas, grade levels, and ability groups (Suskie 258-259). For instance, even though the language usage of grade 12 economics students was hands down the strongest, that skill did not help them outdo the grade 10 AP English students in content mastery and learning strategies. Indeed, economics students' overall dispersion compared more readily to zoology and special education reading.

AP English dispersions seem less of an anomaly when separated out over the two years, in Table 5. Very similar central tendencies and percentages of dispersion emerged:

Table 5. Comparison of Scores/Percentages for 1st and 2nd-Year AP English Folders

Student Learning Outcomes	Mean	S.D.	Exceeds	Meets	Misses
1 st YEAR: AP English (Grade 10, n=20, 100 readings)					
Comprehension of task	2.42	.50	42%	58%	0%
Content	2.34	.57	39%	56%	5%
Strategies	2.52	.52	53%	46%	1%
Language usage	2.55	.50	55%	45%	0%
OVERALL	2.48	.39	20%	53%	27%
2 nd YEAR: AP English (Grade 10, n=16, 48 readings)					
Comprehension of task	2.44	.54	46%	52%	2%
Content	2.40	.54	42%	57%	1%
Strategies	2.46	.62	52%	42%	6%
Language usage	2.56	.50	56%	44%	0%
OVERALL	2.48	.43	21%	52%	27%

What stands out at once is the near-exact overall dispersion of percentages for both years, along with the large percentage of folders that seemed to miss expectations. Means in AP English were so high that when analyzed apart from other subject areas, folders receiving overall scores of 2, “meets expectations,” fell into the category of “misses” instead.

Unlike AP English students, Table 6, below, shows that special education readers in the first year seemed to have stronger means than those in the second year—even though the second-year students seemed to have put in a stronger performance overall:

Table 6. Comparison of Scores/Percentages for 1st and 2nd-Year Special Ed Reading Folders

Student Learning Outcomes	Mean	S.D.	Exceeds	Meets	Misses
1 st YEAR: Sp. Ed. Reading (Grade 9, n=9, 45 readings)					
Comprehension of task	1.93	.61	16%	62%	22%
Content	1.69	.67	11%	47%	42%
Strategies	1.78	.56	7%	64%	29%
Language usage	1.69	.51	2%	65%	33%
Overall	1.79	.47	18%	60%	22%

Student Learning Outcomes	Mean	S.D.	Exceeds	Meets	Misses
2 nd YEAR: Sp. Ed. Reading (Grade 9, n=6, 18 readings)					
Comprehension of task	1.94	.54	11%	72%	17%
Content	1.56	.51	0%	56%	44%
Strategies	1.61	.5	0%	61%	39%
Language usage	1.5	.51	0%	50%	50%
Overall	1.67	.31	22%	67%	11%

Closer examination of the two sets of scores suggest that even with weaker language usage, the year-two special education students comprehended the writing tasks better. However, they appeared to be less proficient at relaying content, and they perhaps found exercising the learning strategies more of a challenge than the year-one students.

Statistical Significance and Meaningful Difference

The pre-/post-writing samples from folder and non-folder groups in this project came from students in separate sections of the same class in each subject area. The folder and non -folder groups were of unequal sizes. For such independent groups, a *t*-test with unequal sample sizes was the appropriate calculation to determine how probable it was that a specific treatment, such as engaging students in writing-to-learn tasks, might result in a difference in the folder groups' outcomes (Steinberg 235; 454).⁶

As with Pearson *r*, the final calculation, *t*, for this two-sample test yields a figure that translates into probability, *p*. For example, in Table 7, *t* for the combined folder and non-folder groups (114) equals 7.581. Probability, *p*, is less than one chance in a thousand that the difference in post-course means between folder and non-folder groups is due to mere chance (see Steinberg 239).

Table 7. Statistical Significance of Score Differences between PRE/POST Writing Exercises*

Subject Areas	Pre-Course Means	S.D.	Post-Course Means	S.D.
COMBINED				
Folder group (n=61)	1.98	.39	2.33	.46
Non-Folder group (n=55)	1.97	.44	1.72	.40

WPA 34.2 (Spring 2011)

Subject Areas	Pre-Course Means	S.D.	Post-Course Means	S.D.
SIGNIFICANT?	No		Yes	
			$t(114) = 7.581, p < .001$	
AP ENGLISH				
Folder group (n=32)	2.11	.38	2.36	.35
Non-Folder group (n=35)	2.17	.35	1.88	.37
SIGNIFICANT?	No		Yes	
			$t(65) = 5.516, p < .001$	
ECONOMICS				
Folder group (n=19)	1.93	.33	2.35	.66
Non-Folder group (n=11)	1.77	.33	1.42	.28
SIGNIFICANT?	No		Yes	
			$t(28) = 4.494, p < .001$	
SP. ED. READING				
Folder group (n=10)	1.59	.30	2.19	.32
Non-Folder group (n=9)	1.41	.33	1.48	.34
SIGNIFICANT?	No		Yes	
			$t(17) = 4.7059, p < .001$	
*Pre/Post data from zoology incomplete.				

Although the folder and non-folder groups scored very similar means on the pre-coursework writing samples, the statistical differences between the folder/non-folder groups on the post-coursework writing samples were consistently significant, suggesting higher learning outcomes in writing for the folder groups. Moreover, calculations for the non-folder groups in AP English and economics show that the means of their post-coursework writing scores even dropped a bit, as did combined groups.

Numbers of pre/post writing samples (61) for the folder groups in Table 7 do not match numbers of folders (88) in Table 5 for several reasons. A few students in the folder groups for AP English and special education reading did not do both a pre- and post-coursework writing sample. Two economics students' folders did not get read during the scoring sessions, but this oversight did not present a reason to exclude scoring their pre/post samples. More unfortunately, data for post-coursework writing from the non-folder

zoology group was not collected, and the pre/post data set for the folder group in zoology was too small to be representative of the class (only six students did pre- and post-coursework writing). Despite the missing data, the theory that writing to learn makes an impact on student learning outcomes strengthens with these statistics.

Yet again, calculating statistical significance for a *t*-test with unequal sample sizes does not suffice because it does not determine the practical importance of an outcome. That is, a calculation to measure effect size had to be performed to determine if the difference between the folder/non-folder groups was in fact not only statistically significant but also large enough to suggest a meaningful impact on common practice (Steinberg 364). The formula for calculating effect size for a two-sample *t* test is called effect size *r*. The acceptable range for a small effect is less than .25; a medium effect, .25 to .40; a large effect, .40 or more (366). Table 8 shows that not only was it highly probable that the writing-to-learn tasks had a statistically significant impact on student learning outcomes, but that the correlation between writing to learn (WTL) and student learning outcomes—in the context of this research—was also consistently large:

Table 8. Effect size *r* of Writing-to-Learn Tasks

Subject Area	Effect size	Correlation of WTL to Outcomes
Combined	.579	Large
AP English	.565	Large
Economics	.647	Large
Sp. Ed. Reading	.752	Large

Table 9 returns to the method of breaking down percentages on how folder and non-folder groups performed on different criteria in the post-coursework writing samples. Although it might have been instructive to separate out all subject areas as well as examine the combined subject areas (as in Table 4), such an extensive analysis did not seem appropriate because of the more informative measures of statistical significance and effect size. A look at the “larger picture,” on the other hand, did seem useful:

Table 9. Scores/Percentages for POST Writing Exercise, Folder/Non-Folder Groups

Student Learning Outcomes	Mean	S.D.	Exceeds	Meets	Misses
FOLDER groups in Econ, AP Eng & Sp. Ed. Reading (n=61)					
Comprehension of task	2.13	.53	21%	71%	8%
Content	2.38	.64	46%	46%	8%
Strategies	2.51	.57	54%	43%	3%
Language usage	2.28	.55	33%	62%	5%
Overall	2.33	.46	28%	59%	13%
NON-FOLDER groups in Econ, AP Eng & Sp. Ed. Reading (n=55)					
Comprehension of task	1.72	.53	4%	65%	31%
Content	1.54	.63	7%	40%	53%
Strategies	1.51	.61	5%	40%	55%
Language usage	1.98	.30	4%	91%	5%
Overall	1.72	.40	11%	65%	24%

While folder groups appear to have benefitted more from writing to learn overall, neither the folder nor the non-folder groups fell more than 5% below expectations in language usage. These descriptive measures suggest the students in the folder group were not necessarily better writers. Yet the difference between dispersion of other scores for folder and non-folder groups was as dramatic as the differences among means. Therefore, although we must exercise caution in making too much of data collected from only one post-coursework writing exercise, it appears that systematically designed write-to-learn tasks did have a positive impact. The folder group showed greater ability to retain course content. The folder group also had a greater command of strategies for thinking critically about that content, whether the students were low-achievers or high, across grade levels 9, 10, and 12.

V. DISCUSSION

In this project, we wondered if quantitative data would support the CSWC's claim that short writing-to-learn tasks not only engage students

in deep learning, but also increase it. Our results promise to contribute to a validation of writing to learn as sound pedagogical practice. However, we discovered an underlying statistical problem: the relation of the number of variables to the number of tests of significance greatly increased the chance that significant effects would be found. Future studies should avoid such a design flaw. Moreover, two aspects of the study's design created limitations that must be addressed: (1) how interrater reliability was established; and (2) how the classes were chosen. Examining these two aspects again leads to insights about method and biases. These limitations can, in turn, become research questions about best practices that may guide future research toward more rigorous controls and, ultimately, toward a more robustly developed, pedagogically useful theory of writing to learn.

This discussion must also focus on a third concern: what happens to research when one of its participants becomes a target of administrative disapproval because of her commitment to a project? Such a concern speaks to the political reasons why quantitative data must become an essential part of our understanding of a pedagogical theory whose roots drink so deeply from humanism and its ethical traditions.

How do we best establish reliability?

The argument that raters need only to decide, essentially, whether a students' collection of written work meets, doesn't meet, or exceeds expectations has existed in our discipline's assessment literature for several years, and it is persuasive (Elbow 1996). Yet if this approach to writing assessment is as reliable as it is persuasive, it should stand up to a measure of co-efficiency such as Pearson r , which checks to see if "the values of accuracy, dependability, stability, consistency, or precision" are "operationalized" between pairs of raters (Parkes 2). The findings indeed suggested as much, but even under carefully established controls and a reasonably well-developed rubric, reliability can be compromised if raters are unsure about assessing student learning outcomes in a subject area different from their own.

Yet again, this project also heeded the caveat that good writing assessment must be balanced with ongoing, thoughtful discussion among the teachers who assess. Only so can a "reading community" evolve "in which reliability grows out of the readers' ability to communicate with each other, to grow closer in terms of the ways they approach the samples" (Hamp-Lyons and Condon 133).

Our qualitative evidence supports this assertion. During the first-year's third-quarter session, teachers rated sample folders which, by then, contained a substantial amount of writing. The zoology teacher remarked:

“Getting to this point takes a long time, but putting everything together and looking at it as a whole was very exciting.” She saw that students “showed consistent writing” and also “showed work that scored consistently” on the rubric. At the end of the year, when it was time to score the entire class set of folders, her colleagues confirmed her remarks. In looking over what students had written, the special education teacher said, “You could actually see the progression of growth through the different animals.” Both the zoology teacher and the special education teacher were struck as well by the similarities between their students. The zoology students were struggling with Latinate terminology that presented as much of a reading challenge to them as the “remedial” material posed to the special education students.

A similar response governed the scoring during the second year. The AP English teacher noted in particular why she scored the special education students’ folders as she did:

The success that stood out most was with [the special education] group. As I read those folders, I struggled. It was a really good struggle, though. The pile of entries that was the smallest would be the one you would think would take the least amount of time. It actually took me the longest. You could see such thought going on in the minds of the students.

The economics teacher agreed, adding, “Often the concerns and successes I had in my room were the same as my colleagues despite the variety of subjects and classroom settings in which we teach.”

The teachers found that meeting regularly throughout the year and talking about the learning objectives they were trying to accomplish with their students in different subject areas ultimately helped them assess the learning outcomes in those subject areas with more confidence. The economics teacher verified as much: “The process allowed us to identify common challenges, successes, strategies, and brainstorm solutions to problems.”

Occasions for such closely controlled inter-disciplinary collaboration are rare, however. Future research—especially in a high school setting—might best be conducted among teachers in the same subject area, where they more uniformly understand the specific content and disciplinary conventions of writing and believe they have the knowledge to assess their students’ mastery of each.

How do we best choose our research samples?

In this project, each teacher chose which class did the writing-to-learn tasks, and which did not. At the very least, this liberty could lead to sample bias because of the likelihood that teachers chose an available folder-group whom they hoped would perform well, or conversely, chose an available non-folder group whom they suspected would perform not so well (see Gay 100). As the project progressed, teachers invested time and effort in developing or revising writing-to-learn tasks, adding to the possibility that experimenter bias could also have increased in concert with their expectations of improved student learning outcomes (219). Moreover, students themselves knew that they were involved in a university project that accounted for the extra writing they were doing, which could have invoked a sense that the folder groups were somehow special—and which could even have caused them to become more interested in (or aloof to) the writing-to-learn tasks.⁷

At the same time, the teachers chose their folder and non-folder classes at the year's beginning when they didn't know the students in either group very well. As the study progressed, developments occurred in the folder groups, causing the teachers to make adjustments that seemed to have an equalizing impact on how they treated both the folder and non-folder classes. For instance, the AP English teacher felt that the students in her year-one folder group sometimes tended to write responses that didn't answer her questions, or more often, didn't demonstrate sufficient use of the learning strategies she wanted them to master. She asked herself: "Why am I doing what I am doing—how am I teaching the material to best fit the needs of my students? Am I accurately measuring their growth and learning?" Consequently, she not only expected students to revise, but she revised her prompts as well. Comparing two of her tasks on *Twelve Angry Men* demonstrates:

Year one:

- Discuss any experiences you have witnessed with our jury system (for example, on television).
- Do you think juries are always fair to the person being tried?

Year two:

- Discuss a time when you worked with a group, explaining trouble, conflicts, satisfaction. Compare it to a situation in the play, where the

jury does or doesn't work together. How do such examples affect your opinion of how people work together?

- After reading the play, discuss why you do or don't think people are truly granted a fair trial. Why do or don't you believe our system works?

Such revisions also affected how she taught her non-folder classes. "I think both classes compare," she observed. "I ended up having the same discussions in both classes. It was as though my prompts guided our class discussions and questions without me even knowing it."

The zoology and economics teachers observed as well that their folder classes informed them how to teach their non-folder classes more effectively. During the first quarter, for instance, the zoology teacher became concerned about her weaker students in the folder class "just copying down information without truly understanding." One student confirmed her suspicions by asking why the class was studying protozoans, why they were looking at these creatures under a microscope, and what protozoans had to do with him. Why indeed? She started to incorporate the kinds of questions he asked, not only into her consequent writing-to-learn tasks, but into her discussions with folder and non-folder classes alike. At the quarter's end, she reported that "most of my students now have an appreciation for the environmental and human connections of the phyla we have studied, including water quality, parasitism, and other dangers to humans." The insight she gained carried through to other units in both the folder and non-folder classes.

In a similar vein, the economics teacher observed: "The students in both classes completed numerous short-answer quizzes as the units progressed and scored very similarly." Moreover, he—along with the student teacher he had the first semester—both noticed that the responses his folder class wrote to the writing-to-learn tasks helped them as teachers develop better test questions and conduct more focused, thorough discussions for the folder and non-folder classes alike.

The economics teacher wanted his students not merely to understand economic principles, but to realize how those principles affected their lives. The writing-to-learn tasks helped him see how he might develop the kind of teacher-student dialogue that could sustain such an objective. He reflected:

On both the final writing prompts and the multiple-choice tests, there was a clear difference in the basic understanding of the units as well as the personal connection that the students [in the folder class] made with the material. I think this outcome is largely due to the

feedback I provided on the prompts, the discussions we had based on the prompts, and the personal insight I gained as a result of reading their written responses to the prompts.

Given the foregoing qualitative data, it seems reasonable to speculate that the participating teachers' concern about their students learning outcomes in both the folder and non-folder classes helped offset some of the biases that this project failed to avoid. And even if the biases still had an effect on the quantitative results, the efforts that the teachers made to create a balance between folder and non-folder classes seem to have helped the results remain meaningful as well. Moreover, the teachers came to appreciate the student outcomes that writing to learn enabled in a variety of academic subject areas, grade levels, and above all, ability levels.

Why do we do research?

Even the best findings in a research project are diminished if they do not have an impact on the context in which the research takes place. This assertion is especially true in a school beset by the everyday problems that accompany high levels of poverty among the students.

During the first year of the project, the special education teacher did not expect that the class she chose to study would present such a major challenge. Black males predominated. As the project moved forward, she suspected that many of the young men had been placed in the class for behavioral issues rather than learning disorders. Accordingly, many of them tended to resist everything she asked them to do—and she had to find ways to work with their resistance. Bit by bit, she began to focus less on the reading program that the district expected her to teach. They hated it. She searched for supplemental materials that she felt were more relevant to her students and deployed the writing-to-learn methods that the program didn't recommend. She knew her students needed more and that state law would back up her efforts (see Quenemoen, et al.).

To illustrate, when the special education students began Mildred Taylor's *The Well: David's Story*, she set the following prompt: "Although it is never said, prejudice plays an important part in *The Well*. Why did Charlie hate the Logans so much?" The students read and reread the relevant passage. Then they discussed the answer. Finally, they wrote. One of her more recalcitrant students turned in a response that said: "Hammer David beat. Knock his ass, No." The special education teacher asked him to read it aloud. He laughed and exclaimed, "What the hell does that say!" She smiled in reply: "You wrote it. You tell me." He revised his answer as follows: "Hammer and David got beat because Hammer hit Charlie. It wasn't

fair, cuz Charlie hit David first. Charlie should have been punished too but he white so he don't get punished"—a concise summary and an accurate interpretation, but one which required extra patience and persistence to elicit.

The high school principal at that time did not support the special education teacher's efforts. Instead, he seemed to hold her responsible when six of the students in her folder group either stopped coming or got expelled, so she had to assign them failing grades. He denied her requests to get involved in school leadership opportunities. He dismissed her expertise in faculty meetings. He cast so many doubts on her teaching abilities and her "reluctance to teach by the book" that at one point, she seriously considered resigning.

Notwithstanding, when the time arrived for the school to do the year's battery of standardized tests, the special education teacher decided that she would do writing-to-learn exercises with her students to help prepare them. She suggested to the principal that the school sponsor a workshop to show other teachers how to do the same. He ignored her. "If he fires me," she said, "maybe I'll go back to school to get my PhD." She redoubled her efforts with her students and did only the most minimal work with the district's required reading program. The result? All of her students who continued to attend class and put forth an effort to do their work managed to pass the tests. At the end of the year, she wrote a kind of contract with herself, drawing up four resolutions:

- Do more writing, even if it's shorter than a paragraph
- Do more scaffolding to get more extended responses
- Create opportunities for more personal connections to the reading
- Don't get sidetracked or discouraged by behavior

In the second year of the project—after the district dismissed the principal—two of her folder-group students who had failed other courses were assigned again to a class of hers for further "remediation." She chose that class as her folder group. She knew they would benefit from another year of writing to learn.

Results from this project may not substantiate the broad claim that special education readers in ninth grade may demonstrate better student learning outcomes if they practice writing to learn. But the results can quantitatively and qualitatively challenge the lacuna in our research that leads critics to say: "For poor writers, or students with little confidence or interest in writing, writing tasks can be detrimental to motivation.... The

content learning of poor writers may not benefit from writing to learn” (Bangert-Drowns, et al. 6; 53). However, to lead teachers away from more narrowly defined classroom assessments that have no connection to valid learning strategies, our research must examine many more “combinations of writing ability and subject area competence,” to produce the “complex interactions” of data that will not only convince teachers, but their administrators as well (6; 53).

VI. CONCLUSION

Collaboration between university writing program administrators and high school teachers most commonly takes place at National Writing Project sites. Nevertheless, educational grants also present an opportunity to pilot productive professional development partnerships. This exploratory project suggests that such partnerships can provide the training and follow-up necessary for cross-curricular faculty to produce statistically significant student learning outcomes through the implementation of writing to learn. This project also suggests the efficacy of specific practices in writing to learn, even in difficult learning environments. By replicating appropriate controls and practices in other secondary and post-secondary settings, future research promises to substantiate a robust pedagogical theory of writing to learn that is solidly based upon quantitative rather than anecdotal evidence. Such quantitative evidence, in turn, can help cross-curricular teachers and writing specialists resist overreliance on standardized tests as the preferred—or oftentimes the only—measure of student learning.

NOTES

1. In Illinois, where the study occurred, public schools rely primarily on local tax bases. Low-income neighborhoods always suffer (Street 54-55).

2. Commentary was very brief, e.g., “Why do you make this claim?”; “What did you leave out of your explanation?” While the reading and English teachers used the general rubric, the zoology and economics teachers substituted descriptions of primary traits for “Content” that more specifically reflected their subject areas—i.e., “understands key concepts in zoology”; “reports accurate economic facts.”

3. Professor Balikrishna Hosmane and graduate assistant Ujjwal Das, checked and provided feedback on this study’s calculations. Professor David Walker, editor of *Multiple Linear Regression Viewpoints*, gave advice on interpreting and presenting results. An anonymous reviewer also offered valuable critique of the “Findings” section.

4. Cohen's Kappa, another measure of interrater reliability, was not used because the rubric did not deploy absolute scoring, e.g., past/fail (see Wood).

5. Our statistical consultants did not suggest calculating effect size here—a measure described below for a t-test—because this use of Pearson r focused on the similarities or differences between readers, not between scores.

6. Pearson r was not calculated for the ratings of the pre/post writing samples between folder and non-folder groups because rating conditions were different. I compared but did not average my own ratings with each teacher of record's. I disagreed with teachers' ratings in 5 instances (3 economics non-folder students; 1 English non-folder student; 1 English folder student). I consistently rated lower overall, and on all disagreements rated lower (e.g. 1.5 to 2.5).

7. This first type of bias among research subjects is known as the "Hawthorne effect," while the second is called a "novelty effect" (Gay 431, 433).

WORKS CITED

- Anderson, Paul, Chris Anson, Bob Gonyea, and Chuck Paine. "A National Study of Writing's Contributions to Learning in College: Major Findings and Practical Implications for All Writing Programs." July 17, 2009. Council of Writing Program Administrators Summer Conference, Minneapolis, MN. Print.
- Anson, Chris. "The Intelligent Design of Writing Programs: Reliance on a Belief or a Future of Evidence." *WPA: Writing Program Administration* 32.1 (Fall 2008): 11- 36. Print.
- Bangert-Drowns, Robert, Marlene Hurley, and Barbara Wilkinson. "The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement: A Meta-Analysis." *Review of Educational Research* 74.1 (Spring 2004): 29-58. Print.
- Bean, John. *Engaging Ideas: The Professor's Guide to Integrating Writing, Critical Thinking, and Active Learning in the Classroom*. San Francisco: Jossey-Bass, 2001. Print.
- Bereiter, Carl and Scardamalia, Marlene. *The Psychology of Written Composition*. Hillsdale: Lawrence Erlbaum, 1987. Print.
- Britton, James. *Prospect and Retrospect: Selected Essays*. Ed. Gordon Pradl. Upper Montclair: Boynton/Cook, 1982. Print.
- Bruner, Jerome. *Toward a Theory of Instruction*. Cambridge: Harvard UP, 1966. Print.
- Duffy, Sean. "Writing in Political Science." *Direct from the Disciplines: Writing Across the Curriculum*. Ed. Mary Segall and Robert Smart. Portsmouth: Boynton/Cook, 2005. 115-24. Print.
- Elbow, Peter. "Writing Assessment in the 21st Century: A Utopian View." *Composition in the Twenty-First Century: Crisis and Change*. Ed. Lynn Bloom, Donald Daiker, and Edward White. Carbondale: Southern Illinois UP, 1996. Print.
- Emig, Janet. "Writing as a Mode of Learning." *College Composition and Communication* 20 (1977): 122-28. Print.

- Flower, Linda, and John Hayes. "The Cognition of Discovery: Defining a Rhetorical Problem." *College Composition and Communication* 31 (1980): 21-32. Print.
- . "A Cognitive Process Theory of Writing." *College Composition and Communication* 32 (1981): 365-87. Print.
- Fulwiler, Toby. "The Argument for Writing Across the Curriculum." *Writing Across the Disciplines: Research into Practice*. Upper Montclair: Boynton/Cook, 1986. 21-32. Print.
- Gay, L. R. *Educational Research: Competencies for Analysis and Application, 2nd Edition*. Columbus: Charles E. Merrill, 1981. Print.
- GreatSchools.net. "Rockford School District 205." <http://www.greatschools.net/city/Rockford/IL>. (28 March 2009). Web.
- Hamp-Lyons, Liz, and William Condon. *Assessing the Portfolio: Principles for Practice, Theory, and Research*. Cresskill: Hampton P, 2000. Print.
- Haswell, Richard. "NCTE/CCCC's Recent War on Scholarship." *Written Communication* 22 (2005): 198-223. Print.
- . (2009, September 8). Re: Assessment. *Writing program administration mailing list*. WPA-L@asu.edu. Archives of WPA-L@asu.edu. <http://lists.asu.edu/archives/wpa-l.htm> (12 Sept. 2009). Web.
- Hesse, Doug. "The Nation Dreams of Teacher Proofing: Neglected Expertise and Needed Writing Research." *College Composition and Communication* 60.1 (2009): W416-W423. <http://www.ncte.org/library/NCTEFiles/Resources/Journals/CCC/0611sep09/CCC0611/Exploring.pdf>. (10 Oct. 2009). Web.
- Jefferson High School Improvement Plan*. Rockford: Jefferson High School, 2005. Print.
- Kozol, Jonathan. *The Shame of the Nation: The Restoration of Apartheid Schooling in America*. New York: Three Rivers P, 2005. Print.
- Klein, Perry. "Reopening Inquiry into Cognitive Processes in Writing-to-Learn." *Educational Psychology Review* 11.3 (1999): 203-70. Print.
- Langer, Judith, and Arthur Applebee. *How Writing Shapes Thinking: A Study of Teaching and Learning*. Urbana: National Council on Teachers of English, 1987. Print.
- McCrinkle, Andrea R., and Carol A. Christensen. "The Impact of Learning Journals on Metacognition and Cognitive Processes and Learning Performance." *Learning and Instruction* 5 (1995): 167-85. Print.
- McLeod, Susan. "Introduction." *Writing Across the Curriculum: A Guide to Developing Programs*. Ed. Susan McLeod and Margot Soven. Newbury Park: Sage Publications, 1992. 1-11. Print.
- Martin, Nancy, Pat D'Arcy, Brian Newton, and Robert Parker. "The Development of Writing Abilities." *Landmark Essays on Writing Across the Curriculum*. Davis: Hermagoras P, 1994. 33-49. Print.
- Morris, Barbara, George Cooper, Constance Childress, Mary Cox, and Patricia Williams. "Collaboration as Sharing Experience: A Detroit Public Schools/University of Michigan Course." *Programs and Practices: Writing Across the Secondary School Curriculum*. Ed. Pamela Farrell-Childers, Anne Ruggles Gere, and Art Young. Portsmouth: Boynton/Cook/Heinemann, 1994. 154-70. Print.

- Nagin, Carl, and National Writing Project. *Because Writing Matters: Improving Student Writing in Our Schools*. San Francisco, CA: Jossey-Bass, 2003. Print.
- National Writing Project. "Research Brief: Writing Project Professional Development for Teachers Yields Gains in Student Writing Achievement." Berkeley: NWP, 2008. www.nwp.org (3 Sept. 2009). Web.
- Newell, G.E. "Learning from Writing in Two Content Areas: A Case Study/Protocol Analysis. *Research in the Teaching of English* 18 (1984): 265-287. Print.
- Newkirk, Thomas. "The Non-narrative Writing of Young Children. *Research in the Teaching of English* 21 (1987):121-44. Print.
- "No Child Left Behind Act of 2001." <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>. (20 June 2010). Web.
- Parkes, Jay. "Reliability as Argument." *Educational Measurement: Issues and Practice* 26.4 (2007): 2-10. Print.
- Prairie State Achievement Examination: Teacher's Handbook 2004-2005*. Springfield, IL: Illinois State Board of Education, 2004. Print.
- "Preliminary Report to be Released to Senior Management and Board Members of the Rockford Public Schools, District 205." Rockford, IL: Hinshaw and Culbertson, LLP, 2007. <http://data.e-rockford.com/upload/files/1/school.pdf>. (3 May 2007). Web.
- Quenemoen, R. F., Lehr, C. A., Thurlow, M. L., Thompson, S. J., & Bolt, S. *Social Promotion and Students with Disabilities: Issues and Challenges in Developing State Policies* (Synthesis Report No. 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. <http://education.umn.edu/NCEO/OnlinePubs/Synthesis34.html>. (7 Aug. 2010). Web.
- "Rockford District 205 Science Curriculum." <http://webs.rps205.com/curriculum/science/>. (12 June 2010). Web.
- Rose, Reginald. *Twelve Angry Men*. New York: Penguin Classics, 2006. Print.
- Steinberg, Wendy. *Statistics Alive!* Los Angeles: Sage, 2008. Print.
- Street, Paul. *Segregated Schools: Educational Apartheid in Post-Civil Rights America*. New York: Routledge, 2005. Print.
- Suskie, Linda. *Assessing Student Learning: A Common Sense Guide, Second Edition*. San Francisco: Jossey-Bass, 2009. Print.
- Taylor, Mildred. *The Well: David's Story*. New York: Puffin, 1998. Print
- Vygotsky, Lev. *Thought and Language*. Trans. Eugenia Hanfmann and Gertrude Vakar. Cambridge: MIT P, 1962. Print.
- Walvoord, Barbara, Linda Hunt, H. Fil Dowling, and Joan McMahon. *In the Long Run: A Study of Faculty in Three Writing-Across-the-Curriculum Programs*. Urbana: National Council of Teachers of English, 1997. Print.
- Wood, James. "Understanding and Computing Cohen's Kappa: A Tutorial. *Web-PsychEmiricist*. Retrieved May 28, 2010 from http://wpe.info/papers_table.html. Web.
- Yancey, Kathleen Blake. *Reflection in the Writing Classroom*. Logan: Utah State UP, 1998. Print.
- Zinsser, William. *Writing to Learn: How to Write—and Think—Clearly about Any Subject at All*. New York: Harper & Row, 1989. Print.