



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Assessing Writing 9 (2004) 56–75

ASSESSING
WRITING

Assessing and teaching what we value: The relationship between college-level writing and critical thinking abilities

William Condon*, Diane Kelly-Riley

Washington State University, Writing Center, Avery 487, Pullman, WA 99164-5046, USA

Abstract

Washington State University (WSU), has developed two large-scale assessment programs to evaluate student learning outcomes. The largest, the Writing Assessment Program, diagnoses student writing abilities at entry and mid-career to determine the type of support needed to navigate the expectations of our writing-rich curriculum. The second, the Critical Thinking Project, has developed an assessment instrument, the *WSU Guide to Rating Critical Thinking*, adaptable by faculty to their instructional and evaluative methodologies, which we can employ across the curriculum to evaluate student critical thinking outcomes. The development of these two measures has provided insights into limitations of each measure and the student learning outcomes produced. Further, the results of our studies question current mainstream writing assessment practices, common assumptions about writing and critical thinking, and several aspects of higher education classroom and curricular praxis.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Assessing; Teaching; College-level writing; Critical thinking

Writing is the coin of the realm here. It permeates the whole atmosphere rather than being compartmentalized into a single course or slapped on as a series of skills. We believe writing is the tool of thinking. The best way to learn to think is to read a lot of good writing and write a lot about what you've read. Writing and the communication of ideas are central to all disciplines whether one is in college or the workplace. One of the most important skills in the digital age is, in fact, one of the oldest — *writing*.

* Corresponding author.

E-mail address: bcondon@wsu.edu (W. Condon).

Michelle Toleia Myers, President, Sarah Lawrence College

Having to think about it, instead of just write it, took some getting used to . . .

Danielle Ozaki, World Civilizations student
(quoted with permission)

While most academic articles proceed on a deductive pattern, the particular set of problems we explore here seem to require an inductive method, by which means we can take readers through our process of discovery and engage in discussions along the way of the various major issues we want to explore. In our own program development efforts, we fell prey to the same assumption about the relationship between writing and critical thinking that is stated in the first epigram, above; yet, as we delved deeper, we came up against the phenomenon that Danielle Ozaki articulates in the second epigram. Exploring the ways in which writing and critical thinking both are and are not necessarily related led us to a closer examination of the Critical Thinking Project we describe here in some detail. Finally, the very fact that we were engaged in — and therefore *could* compare — complex, robust assessments of student learning outcomes led us to a set of questions about some of our own assessment practices, which mirror commonly accepted practices for the direct testing of writing. Since these challenges arose and came to exhibit their connections over half a decade of various investigations, it seems best to explore them more or less chronologically, stopping along the way to unpack the issues that arise.

1. Program history

Beginning in 1987, Washington State University fostered a comprehensive reform of its General Education program. Early in this process, we defined a set of objectives for the baccalaureate degree, an integrated set of writing requirements, and an integrated structure spanning students' academic careers. Through these changes we hoped to transform the typical "smorgasbord" curriculum into a coherent learning experience attentive to the broader competencies of writing and higher order thinking skills. Writing instruction and practice were integrated at every level and within every discipline as the requirements extended into the junior and senior year. At two points, entry and mid-career, students' writing was evaluated diagnostically to identify whether they needed additional support as they proceeded through the lower- and upper-division writing requirements (Haswell, 2001).

General Education reform and the development of WSU's Campus Writing Programs proceeded, in part, in response to a state-imposed mandate for entry-level, mid-career, and end-of-program assessments. Throughout the implementation of the Writing Program, we believed, like most others, that writing and critical thinking were inextricably linked: as the first of our epigrams indicates, a common assumption holds that to improve students' writing is necessarily to improve their abilities as thinkers. As our students wrote their ways through Zoology labs, International Business case studies, term papers, exam questions, textual analyses,

lesson plans, and so on, we were certain that their intellectual abilities improved as well.

In the late 1990s, our beliefs changed about the relationship between writing and critical thinking. We could internally document improvement in student writing by tracking student progress within our Writing Program. Other studies also documented student growth within our system and their growth as writers (Haswell, 2000). The 2001 Progress Report on the Writing Portfolio showed that 90% of student writers received passing ratings or higher on junior-level writing portfolios, indicating an overwhelming majority of upper-division students demonstrating writing proficiency as defined by WSU faculty (Burke & Kelly-Riley, 2002).

Our faculty, though, lamented that students lacked adequate higher order thinking abilities — a sentiment echoed by many faculty who evaluated our junior Writing Portfolio — so we began more systematically exploring the relationship between writing and critical thinking. In 1999, using an earlier version of the *WSU Guide to Rating Critical Thinking* (see Appendix A), which we had first developed in 1997, we evaluated papers for critical thinking written for three different senior-level capstone courses. Surprisingly, they revealed low critical thinking abilities (a mean of 2.3 on a six-point scale). This phenomenon, in which writing deemed acceptable in quality despite lacking evidence of analytic skills, was also discerned among other lower division General Education courses. In one workshop, 25 instructors of the World Civilizations core course evaluated a freshman paper in two ways — in terms of the grade they would give (they agreed on a B range) and in terms of critical thinking (a score of 2 on a six-point scale). The conclusion they arrived at informally was that as an instructor group, they tended to be satisfied with accurate information retrieval and summary and did not actively elicit thinking skills in their assignments. These forays led us to suspect that in education praxis there may often be little, if any, relationship between writing and critical thinking. Courses that are designed to promote, among other abilities, higher order thinking, and which are taught by faculty who believe that they are in fact eliciting those abilities, nevertheless fail to do so. The fact that writing was the primary vehicle, in our General Education Program, for promoting these competencies gave us the first inkling that no automatic connection between writing and critical thinking exists, even in curricula and classrooms where the two are explicitly linked.

At this time, our state legislature, like many others, threatened to institute state-wide accountability measures for publicly funded institutions of higher education. Anticipating a state-mandated measure for critical thinking, and pursuing our own desire to develop an instructionally useful assessment tool, faculty from the Center for Teaching, Learning and Technology (CTLT), the General Education Program, and the Writing Programs collaborated to develop the *Washington State University Guide to Rating Critical Thinking*. This *Guide* was derived from scholarly work, including Toulmin (1958), Paul (1990), Facione (1990), and local practice and expertise, to provide a process for improving and a means for measuring students' higher order thinking skills during the course of their college careers. Our intent was to develop a fine-grained diagnostic of student progress

as well as to provide a means for faculty to reflect upon and revise their own instructional goals, assessments, and teaching strategies. The *Guide* can be adapted instructionally and can be used as an evaluative tool — applying a six-point scale for evaluation and combining ETS scoring methodology with expert-rater methodology (Haswell, 1998a, 1998b; Haswell & Wyche, 1996). The resulting *WSU Guide to Rating Critical Thinking* identifies seven key areas of critical thinking:

- identification of a problem or issue
- establishment of a clear perspective on the issue
- recognition of alternative perspectives
- location of the issue within an appropriate context(s)
- identification and evaluation of evidence
- recognition of fundamental assumptions implicit or stated by the representation of an issue
- assessment of implications and potential conclusions

A fully developed process or skill set for thinking critically demonstrates competence with and integration of all of these components of formal, critical analysis.

In December 1999, we more formally explored the relationship between writing and critical thinking as demonstrated in the WSU Writing Assessment Program. Our assessments — both of writing and of critical thinking — define the constructs operationally. In the Writing Assessment Program, students' writing samples are evaluated by teachers from the courses *into which* the students will be placed (Haswell, 1998a, 1998b, 2001; Haswell & Wyche, 1996). Teachers read two timed writing samples, one analytical and the other reflective, and concentrate on the criteria of Focus, Organization, Support, Fluency, and Mechanics. Using these categorical criteria, teachers are asked (1) to compare the writing in the samples to the kinds of writing their successful students produce in order to decide whether the students producing the samples are ready for the course(s) and (2) to consult with each other over difficult cases. Faculty define “good writing” in the course of making their decisions, and the assessment procedures help faculty maintain consistency over time in those decisions (cf. Smith, 1993 for similar results from a similar system). This expert-rater system has been widely cited for its reliability and its potential for linking assessment with instruction (see Huot, 1996, e.g.), and its context-responsiveness (see, e.g., Elbow, 1994). It is also a highly reliable scoring procedure, demonstrating scoring outcomes that are consistent at as much as a 98% rate (Haswell, 1998a, 2001).

Similarly, the *WSU Guide to Rating Critical Thinking* acts as a description rather than a definition of critical thinking. Only in the process of rating samples do faculty operationalize the *Guide*'s dimensions into something like a definition of the construct. Again, the rating procedures ensure that faculty rate thoughtfully and consistently so that the operational definition remains constant over time. Using a

six-point scale for each dimension, faculty select one of the following levels:

1. Not evident; can't find it anywhere in the paper.
2. Discernable, but not developed.
3. Better than 2, but not yet 4. Could be confused, inconsistent, etc.
4. Important to the paper.
5. Better than 4, but not yet 6. May be substantially developed in places, but not throughout the paper.
6. Substantially developed; considered in full complexity; nuanced and sophisticated.

The system results in reliability coefficients of .8 and higher (often much higher), primarily because not only do faculty focus on one dimension at a time, but they are, in effect, using a low-medium-high (2, 4, 6) score set, with the odd numbers acting as an “in between” score.

In the WSU Writing Assessment Program, student writing is diagnostically evaluated at entry and at mid-career to identify students requiring additional writing support. We compared these results to an evaluation applying the critical thinking evaluation methodologies to the same writing samples. Furthermore, we compared data collected from courses specifically designated to integrate the *Guide* into their evaluative and instructional methods with courses that did not. These initial studies yielded interesting results. First, by examining the freshman- and junior-level timed writing samples, we found that students' critical thinking abilities increased by a statistically significant amount between freshman and junior years without any exposure to overtly stated expectations regarding higher order thinking. The prompt itself, of course, does indicate that student's performances must include the kinds of higher order thinking that are included on the *Guide*. Students writing at entry-level received a mean critical thinking score of 2.59 ($SD = 0.738$), while mean junior-level critical thinking scores increased significantly to 3.05 ($SD = 0.791$) ($P = .001$). The 0.458 overall gain reflects significant differences in performance on all seven dimensions of critical thinking identified in the *Guide*, yet the mean of 3.05 is still only halfway up the six-point scale.

We also ascertained that students' critical thinking scores improved more in one semester in courses that overtly integrated the *Guide* than from the regular progression from freshman to junior year, as established by performances in WSU's Writing Assessment Program. Students in courses in which the *Guide* was overtly used increased their scores up to three and a half times as much as students in courses that did not. In one instance, papers were rated from two different semesters of Entomology 401, Biological Thought and Invertebrates, representing a single course and instructor, one semester when the *Guide* was not used ($n = 14$), and from the following semester when the *Guide* was used ($n = 12$). The overall mean score in the semester without the *Guide*, 1.867 ($SD = 0.458$), increased significantly to 3.48 ($SD = 0.923$, $P = .001$) the semester when the *Guide* was used. In the semester when the *Guide* was not used, the instructor read and commented on drafts. The

following semester, the instructor adapted the *Guide* into a feedback form targeted to five of the seven dimensions. The instructor met with each student to talk about the drafts in terms of the identified areas of evaluation (see [Appendix B](#)).

Gains in critical thinking were further supported in studies observing courses that implemented the *Guide* as opposed to courses that did not. One hundred twenty-three student essays from several lower and upper-division undergraduate courses were assessed for critical thinking. In the four courses where the *Guide* was used in varying ways for instruction and evaluation ($n = 87$), the papers received significantly higher critical thinking ratings than in the four courses in which the *Guide* was not used ($n = 36$). The mean score for courses in which the *Guide* was not used was 2.44 ($SD = 0.595$) compared to 3.30 ($SD = 0.599$, $P = .001$) in courses that employed the *Guide*.

The most surprising revelation in these early studies was an inverse relationship between our scoring of student work in our Writing Assessment Program — the entry-level Writing Placement Exam and the junior-level timed writing portion of the Writing Portfolio — and our evaluation of the same work using the *WSU Guide to Rating Critical Thinking*. In other words, the better the writing, the lower the critical thinking score, but the more problematic the writing, the higher the critical thinking score. Sixty samples of writing, representing pairs of entry-level Writing Placement Exams and junior-level timed writing portions of the WSU Writing Portfolio, were evaluated using the *WSU Guide to Rating Critical Thinking* in order to gather general baseline data regarding the critical thinking abilities of students at WSU. This population represented students who wrote on topics that required them to analyze a subject, but who had no prior exposure to the *Guide*. Students deemed better prepared for the rigors of academic writing by the freshman level Writing Placement Exam had lower critical thinking scores at a statistically significant level ($r = -.339$, $P = .015$). The same inverse correlation phenomenon appeared in the rating of the junior-level timed writings, though the results were not statistically significant ($r = -.169$, $P = .235$). It seemed that our own writing assessment practice tended to elicit and reward surface features of student performance at the expense of higher order thinking.

Recently, this study was replicated with 20 paired samples of student writing from the Writing Placement Exam and the Writing Portfolio. The timed portion and the three course papers from each student's Portfolio were included. The students who produced these samples of writing had not been exposed to the *WSU Guide to Rating Critical Thinking*. In the second study, we wanted to see whether the inverse correlation continued and how the course papers in the Writing Portfolios would perform for a critical thinking evaluation. Furthermore, we re-evaluated the pairs of Writing Placement Exams and Portfolio timed writings with a nine-point critical thinking scale (rather than the regular six-point scale) allowing for more discrete analysis of writing and critical thinking. To do this, we broke each of our usual three levels of scoring into three, yielding nine score levels (see [Table 1](#)). This, we hoped, would allow us to make finer discriminations, so that we could better understand the results, particularly if the inverse correlation held up.

Table 1
Interpolation of writing placement scores to nine-point scale

CT score	1	2	3	4	5	6	7	8	9
Writing Placement Exam	Low, English 100	Mid, English 100	High, English 100	Low, English 101 + 102	Mid, English 101 + 102	High, English 101 + 102	Low, English 101	Mid, English 101	High, English 101
Writing Portfolio	Low, needs work	Mid, needs work	High, needs work	Low, acceptable	Mid, acceptable	High, acceptable	Low, exceptional	Mid, exceptional	High, exceptional

At the Writing Placement level, English 101 represents the regular first-year composition course placements; English 101 + 102 represents the regular first-year composition course requiring supplemental tutorial instruction; and English 100 represents basic writing. At the Writing Portfolio level, exceptional represents the top 10% of writers; needs work represents 10% of writers requiring supplemental upper-division writing support; and acceptable represents the 80% of writers who are ready to meet the challenges of upper-division writing requirements.

In this analysis, no statistically significant critical thinking growth was observed between freshman and junior years either between the timed samples or between the timed freshman exam and the revised work included in the junior Writing Portfolio. The mean freshman level critical thinking score was 3.07 ($SD = 0.97$), and it increased to 3.17 ($SD = 0.89$) for the portfolio timed writing and a mean score of 3.21 ($SD = 0.61$) for the average critical thinking score for the papers in the Writing Portfolio. Impromptu writing did not yield strong higher order thinking responses, nor, surprisingly did course assignments. A regression analysis concluded that no relationship existed between our writing assessment scores and our critical thinking scores.

Our findings demonstrate the separate nature of writing and critical thinking. Writing professionals have held the belief that writing and critical thinking are inextricably linked — often enough, as in our first epigram, the two are simply equated. Early essays in the field of composition and rhetoric established this long held assumption. Emig (1977) argues, “some of the most distinguished contemporary psychologists have at least implied such a role for writing as a heuristic . . . [They] have pointed out that higher cognitive functions, such as analysis and synthesis, seem to develop most fully only with the support system of verbal language — particularly it seems, of written language” (p. 122). Emig notes the implication of the relationship between writing and critical thinking by such theorists as Vygotsky, Luria, and Bruner. For Writing Across the Curriculum programs, McLeod (1992) argues that “writing is not only a way of showing what one has learned but is itself a mode of learning — that writing can be used as a tool for, as well as a test of, learning” (p. 4).

Both constructs — writing and critical thinking — are abstract, complex, socially constructed, contextually situated terms, and this presents problems in analyzing our conflicting results. Anyone trying to specify what makes up “good writing” faces a daunting task, since that construct differs widely from discipline to discipline and from context to context. Good writing in a history class — narrative-based argument, say — is problematic even in another Humanities discipline, English Studies, where narrative-based arguments are neither highly valued nor widely practiced. Needless to say, “good writing” from History or English would likely be considered *bad* writing in most science classes. The same problem faces the task of defining critical thinking: the type of critical thought required by a student in a Turf Management or Orchard Management course would be vastly different from the type of reasoning used in a Metaphysics Philosophy

course. The kind of critical thinking is driven by the values and the types of work required in the discipline.

2. What is critical thinking anyway?

The current literature on critical thinking is rife with conflict and competing ideologies. Paul, Elder, and Bartell (1997) defined critical thinking in their study on faculty knowledge as “thinking that explicitly aims at well-founded judgment and hence utilizes appropriate evaluative standards in the attempt to determine the true worth, merit, or value of something.” Halpern (1997) asserts that critical thinking is the “use of those cognitive skills or strategies that increase the probability of a desirable outcome. It is used to describe thinking that is purposeful, reasoned and goal directed” (p. 4). From a cognitive psychologist’s view, she cites several other definitions from that perspective: critical thinking is the “formation of logical inferences”; it is the development of cohesive and logical reasoning and patterns; it is careful and deliberate determination of whether to accept, reject or suspend judgment; it is mental activity useful for a particular cognitive task (1997, p. 4). Facione (1990) asserts that critical thinking is “purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based.” Scriven and Paul (2003) define critical thinking as “the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing, and/or evaluating information gathered from, or generated by observation, experience, reflection, reasoning, or communication as a rubric to belief and action.”

Additionally, research has been done regarding the development of college-level students, and these theories are often included in the views regarding students’ cognitive growth. Haswell (1991) describes this growth of writers as a process of alienation and reconciliation. Perry (1968, 1981) charts students’ growth through various and increasingly complex stages, and college educators often cite Lawrence Kohlberg’s Stages of Moral Development (Kohlberg, 1984) as a way to categorize students’ development in college. While these theories do not directly define critical thinking, they are looked to as a way to describe the processes of maturation that include cognitive abilities. Regardless of these descriptions, however, Paul et al. (1997) assert that instructional faculty largely do not know how to describe or define critical thinking beyond trendy pedagogical buzz words even though “the vast majority (89%) stated that critical thinking was of primary importance to their instruction.”

Rather than attempting to create an all-encompassing definition of critical thinking, the Washington State University Critical Thinking Project encourages faculty to create contextually based definitions and applications of critical thinking. Our intent is to use the *Guide* as a diagnostic measure for student progress, and to provide faculty a means to reflect upon and revise their practices. No one definition

of critical thinking is applicable to every discipline at every level. Faculty improve and assess the critical thinking abilities of their students by the improvement of their practice as teachers. This includes demystifying the expectations they have for students, articulating their values for the classroom and the types of disciplinary abilities they want to emphasize. In the development of the *Guide*, we relied largely upon the expertise of faculty to inform us what a tool eliciting critical thinking might contain. We came up with a list of seven dimensions that seemed generally to comprise features of critical thinking across disciplines. Teachers are encouraged to use as many or as few of the seven to emphasize in their classrooms based on the discipline, their teaching styles, the makeup of the students in their course, and so on.

Faculty are encouraged to take the seven-dimension *Guide* and create evaluation criteria and assignments that suit their instructional styles and disciplinary expectations. They create criteria to evaluate student work. Most distribute these criteria to students before assignments so that students can develop a clear understanding of expectations. We also encourage faculty to design assignments that promote higher order thinking. We have found that without a tool such as the *Guide*, faculty predominantly develop assignments based on information retrieval — assignments that do not give students the opportunity to apply or engage with the course material and so counteract their evaluation criteria. Most of the participants in our Project have developed criteria and assignments dealing specifically with writing, although a few are working to develop non-written assignments and evaluation criteria in disciplines such as mathematics, physics, crop and soils sciences, and web design.

Given an interaction between two such complicated constructs, the results we obtained are really not unexpected. Our job in this article is to explore those results, to explore their meaning, and to start identifying ways in which we can improve our practice of assessing, promoting and teaching writing. Given the context we have described above, then, we can begin to draw some conclusions about widely accepted assumptions regarding critical thinking, writing across the curriculum and writing assessment — and in the process, raise more questions than we can, at this point, draw conclusions. Still, at this point we can point to improvements we can make in both assessment and instruction.

3. Does the inverse correlation mean we've had it all wrong?

Murphy and Ruth (1993) argue “we need to consider the adequacy of traditional psychometric field-testing procedures for auditing and appraising the interactions of examinees with topics in writing assessments” (p. 267). The Critical Thinking Project shed light on limitations in our Writing Assessment Program, and in our efforts to promote Writing Across the Curriculum. The inverse correlation, and then the lack of relationship between our writing assessment scores and critical thinking scores point to what anecdotal evidence has long supported. Oftentimes,

raters in our Writing Assessment Program comment that the exams seem to show sound writing abilities, but really contain no critical thought, or are vacuous or superficial. Haswell's research (1991) indicates that when writers take risks with new ways of thinking, often their writing breaks down in structure as the student grapples with a new way of thinking.

These assumptions have led us to reconsider the relationship between writing and critical thinking, and how they play out in large-scale assessment programs situated and defined by local context. The lack of relation does not mean that either assessment is patently wrong. The lack of relationship between writing scores and critical thinking scores indicates that having students write does not automatically mean that we ask students to think critically. This point is surprising for many writing professionals because we have operated with the assumption that writing and thinking are inextricably linked. Theorists like Emig and McLeod — as well as the U.S. Government's Office for Educational Research and Innovation ([National Center for Education Statistics, 1993, 1994, 1995](#)) — assert a direct connection. In addition, research into the relationships between cognitive abilities confirms that, at some level, any two cognitive abilities are related. Arthur Jensen (1994), e.g., asserts, "I have found no evidence of any two or more mental abilities that are consistently uncorrelated or negatively correlated in a large unrestricted sample of the population." Such studies, however, not only depend on acontextualized measures of cognitive abilities — as opposed to measures that examine student learning outcomes — but they also measure "ability in the abstract" (Conrad, 1989). Students, not being Lord Jim, are being asked to perform highly developed, advanced, *learned* competencies: not merely critical thinking, but college-level critical thinking; not merely writing, but college-level writing in the disciplines. So one problem with the common assumption that equates writing with critical thinking is that so much depends on the context surrounding the performance and the method for measurement. In our context, of course, the reasons for the separation we found in practice should be fairly clear:

1. If faculty do not explicitly ask for critical thinking, students do not feel moved to do it;
2. If faculty do not define the construct *critical thinking* for students, students will not produce a definition;
3. If writing tasks call for summary and fact reporting, we have no reason to suspect that students' performances will incorporate critical thinking;
4. If faculty do not receive assistance in developing assignments that set high expectations and that explain clearly what those expectations are, there can be no reason to assume that course assignments and materials will include either.

Writing acts as a *vehicle* for critical thinking, but writing is not itself critical thinking. The inverse correlation points to the need we have, as writing

professionals, to acknowledge the possibilities and limitations for writing instruction and evaluation practices. We are getting both good news and bad news from our results — that is, the assessment is showing us where our practices are working and where we need to improve. Flawed assessments might yield the illusion of one of those outcomes, but not all together. We need to consider how overtly integrating critical thinking expectations into our writing instruction, writing assessment and faculty development practices can complement the work we already do.

4. The problem with timed exams

The non-relationship between writing and critical thinking questions the role of holistically scored timed writings — the most widespread method for the direct testing of writing. To this point, our data clearly indicate the problem — the disconnect — but not a full explanation. So far, we have made the following observations.

The nature of the timed sample undervalues higher order thinking in the construct we are testing. Our writing prompts ask students to read a short passage, to analyze the author’s position(s), and to respond in a variety of ways — to agree or disagree, to suggest a better solution, to reframe the issue in a different context, and so on. In other words, these prompts explicitly request responses that engage students in the same dimensions of critical thinking included in the *Guide*. Yet raters uniformly complained that the samples rarely include “thinking.”

In the WSU timed writing assessments, students have two hours to produce two essays: a longer, analytical or argumentative piece and a shorter, reflective one. Students are advised to spend an hour on the longer essay, a half hour on the reflective piece, and a half hour looking back over their work and revising where necessary. In timed writing terms, this amount of time is on the generous side (most of ETS’s timed writings allow 20–30 minutes); the inclusion of two samples of different genres should allow students to show more of their abilities. In other words, as timed writings go, this one provides as much opportunity to demonstrate critical thinking as students are likely to get in any assessment based on timed writing.

Likewise, the fact that college juniors, on an assessment where they have a significant stake, barely achieve an average score above three on a six-point scale — while at the same time demonstrating writing competencies that faculty rate as (at least) competent — suggests a flaw in the nature of timed writings. If we are trying to test for thinking abilities, and not merely for the ability to produce a short sample of basically correct prose, then the timed writing may not fulfill our needs. The limitations of time — perhaps *any* unnatural limitation on time — and the fact that students are long trained by various educational assessments in their K-12 schooling to consider timed writing in various reductive ways may indicate

that the time has come to retire the timed writing method for direct testing of writing in all cases that ask students to demonstrate broader sets of competencies. In other words, timed writings scored via primary trait scales for limited abilities — say, the production of mechanically correct sentences, the idiomatic use of a language, etc. — may still be valid, but using timed samples for any larger purpose — or to assess any more complex set of competencies — is suspect, at the very least.

Holistic rating scales provide a false picture of student performance. One significant difference between holistic scoring and the *WSU Guide to Rating Critical Thinking* is that a writing sample receives a single score — based on the rater's overall impression — from a holistic method, whereas the *Guide* provides seven separate scores, one for each dimension. And since the development process involved a determination that the dimensions were actually rating separate, non-correlating aspects, each sample receives seven *different* scores. Haswell (1991) notes that holistic scoring tends to flatten a student's performance, to bring up the low points and undervalue the accomplishments. In short, a holistic score provides the basis for a rough ranking and nothing more. That is, the student receiving a 4 is somehow a better writer than a student who receives a 3, though the differences *among* 3s or 4s may be greater than the differences between a specific student's 3 and another's specific 4. Raters collapse a wide range of specific judgments into one overall impression; in that act, they conceal a considerable variety in a writer's strengths and needs.

The *Guide* does not collapse these judgments. It leaves them separate — except in the research process of averaging the scores, and the disappointingly low average scores perhaps confirm the very phenomenon we wonder about in the holistic score.

The amount of instructionally useful feedback generated from a holistically scored sample is limited. In most cases, a holistic score is not intended to provide much if any information that a teacher might use to plan instruction for a given student or set of students. All a student or a teacher sees is a score report, a rough means of ranking one student above or below the next. Even in our Expert-rater system, where teachers from the destination courses perform the rating, the scores provide little to no useful feedback. The act of rating provided teachers with useful information about the range of abilities in the student population as a whole, but the limited nature of the writing task and the time constraints involved in producing the sample tend to limit teachers' abilities to rely on the sample as a valid test of any given student's true writing abilities.

By contrast, the *WSU Guide to Rating Critical Thinking* allows us to score actual student learning outcomes, and it provides a more fine-grained description of a student's abilities, a description that reveals strengths and weaknesses. The *Guide* also serves instructional purposes. Since the construct *critical thinking* is situated in disciplinary contexts, adapting the *Guide* for their own courses prompts faculty to define critical thinking within their own disciplines. Since faculty share that adaptation with students, the *Guide* serves to help faculty communicate their expectations

to students. And since faculty incorporate the adapted *Guide*'s language into their assignments, those assignments become more likely to elicit performances of critical thinking. Once we are able to provide finer grained assessments that act to improve instruction, will we ever want to go back to a holistically scored timed sample?

5. On the road to improving practice

Writing Across the Curriculum practices promote writing instruction as a shared commitment across campus and introduce strategies to promote writing within a disciplinary context. No one department is responsible for the education of writers. Each instructor in every discipline has the responsibility of teaching students the conventions and expectations of their particular domain. Our collective efforts lead to better prepared writers. Our results indicate, however, that we need to examine these practices, and not assume that the learning-to-write and writing-to-learn strategies will necessarily take care of the problem of student thinking. The results of our studies can help inform ways in which we can improve practices in our field.

First, our results provide two distinct lenses for viewing student work. Multiple measures within robust assessment systems yield a more complicated portrait of what faculty teach and what students learn. This complexity allows us to make choices and decisions about the instructional directions we wish to take. Efficiency in some assessment models serves important purposes: students need to take entry-level courses, and they need to be matched up with the best possible instruction. However, this match or rank can only give so much information, and other forms of assessment can yield complementary information about our practices too.

Second, writing has an important place in higher education, but it is only one vehicle through which many educational objectives can be achieved. Writing is one way we can promote abilities and competencies valued in various disciplines. Critical thinking is a value that all disciplines want to promote, and it can be promoted through writing, but such promotion needs to be done overtly. We don't have any easy answers on ways to promote critical thinking through writing. We're collectively trying to figure that out. We can say that it is a process through which we need to articulate our values within the important contexts in which we reside, and ensure that our praxis matches our expectations. Our faculty's uses of the *Guide* indicate that it — and the faculty development process that has accompanied it — provides an effective response to at least part of the challenge.

Third, our findings point to the possibilities for working with student learning outcomes. Ed White (1996) said that how we assess depends on what we want as a result of the assessment. If all we want to do is rank students, then our current procedures serve us well enough. But if we want to compile fuller senses

of our students' abilities, then we need to pursue better assessment mechanisms. If we want to be able to use the results of our assessments as a basis for instruction, then we need assessments that yield more, not less information. These assessment mechanisms should be located directly within classroom contexts and should look at actual products from students rather than from independent, separate testing occasions. We should not say to our students, "Stop learning now so that we can take a test." Currently, there is much energy and thought expended on ways to create better one-shot timed exams — from getting high rater agreement between computers and humans to providing a computerized environment to take a multiple choice test on English to place students into local classroom environments. The tests tell students that we don't really care about what they can do. We just want placing them to be as easy as possible, and to require as little effort and human contact as possible. These are the wrong messages. Students' rapid improvement in classes where the *Guide* was centrally employed show us what can happen when faculty provide clear instructions and evaluation criteria and students' learning outcomes are the focus of evaluating the students' performances, whether in the form of a grade or a broader, institutionally oriented evaluation.

In conclusion, our studies, beginning with the inverse correlation and proceeding through the rest described in this article, reveal a need to look seriously at educational praxis in higher education, first to be sure that we actually promote the values and competencies we claim to promote, and second to be sure that any assessment that purports to identify those values and competencies actually does so. The failure of our writing assessments to identify the full set of student competencies we thought we were testing for is only one indication that the so-called direct testing of writing, at least in the form of timed writings, is far less direct than we need it to be. In the context of higher education (and we would extend the same statement to secondary education, though we have no data to establish it in that context) the set of values and competencies we commonly attempt to promote in our students is extremely complex. In addition, all these values and competencies, like critical thinking, are socially constructed and highly situated within different disciplines. In order to evaluate our students' performance, we need assessments that use the learning outcomes from classes, which means that we must develop assessment tools and processes that are capable of evaluating those outcomes.

Acknowledgments

The authors would like to thank Dr. Carol Sheppard, Department of Entomology, Washington State University, for the use of the adaptation of her rubric from Entomology 401. We would also like to thank the Council of Writing Program Administrators for a grant award which supported the research in this article.

Appendix A

Guide to Rating Critical Thinking
Washington State University
2001

- 1) Identifies and summarizes the **problem/question at issue** (and/or the source's position).

Scant

Substantially Developed

Does not identify and summarize the problem, is confused or identifies a different and inappropriate problem.	Identifies the main problem and subsidiary, embedded, or implicit aspects of the problem, and identifies them clearly, addressing their relationships to each other.
Does not identify or is confused by the issue, or represents the issue inaccurately.	Identifies not only the basics of the issue, but recognizes nuances of the issue.

- 2) Identifies and presents the **STUDENT'S OWN perspective and position** as it is important to the analysis of the issue.

Scant

Substantially Developed

Addresses a single source or view of the argument and fails to clarify the established or presented position relative to one's own. Fails to establish other critical distinctions.	Identifies, appropriately, one's own position on the issue, drawing support from experience, and information not available from <i>assigned</i> sources.
---	--

- 3) Identifies and considers **OTHER salient perspectives and positions** that are important to the analysis of the issue.

Scant

Substantially Developed

Deals only with a single perspective and fails to discuss other possible perspectives, especially those salient to the issue.	Addresses perspectives noted previously, and additional diverse perspectives drawn from outside information.
---	--

- 4) Identifies and assesses the key **assumptions**.

Scant

Substantially Developed

Does not surface the assumptions and ethical issues that underlie the issue, or does so superficially.	Identifies and addresses the validity of the key assumptions and ethical dimensions that underlie the issue.
--	--

5) Identifies and assesses the **quality of supporting data/evidence** and provides additional data/evidence related to the issue.

Scant

Substantially Developed

Merely repeats information provided, taking it as truth, or denies evidence without adequate justification.	Examines the evidence and source of evidence; questions its accuracy, precision, relevance, completeness.
Confuses associations and correlations with cause and effect.	Observes cause and effect and addresses existing or potential consequences
Does not distinguish between fact, opinion, and value judgments.	Clearly distinguishes between fact, opinion, & acknowledges value judgments.

6) Identifies and considers the influence of the **context** * on the issue.

Scant

Substantially Developed

Discusses the problem only in egocentric or sociocentric terms. Does not present the problem as having connections to other contexts—cultural, political, etc.	Analyzes the issue with a clear sense of scope and context, including an assessment of the audience of the analysis. Considers other pertinent contexts.
--	---

7) Identifies and assesses **conclusions, implications and consequences.**

Scant

Substantially Developed

Fails to identify conclusions, implications, and consequences of the issue or the key relationships between the other elements of the problem, such as context, implications, assumptions, or data and evidence.	Identifies and discusses conclusions, implications, and consequences considering context, assumptions, data, and evidence. Objectively reflects upon the their own assertions.
--	--

Contexts for consideration

Cultural/social Group, national, ethnic behavior/attitude	Scientific Conceptual, basic science, scientific method
Educational Schooling, formal training	Economic Trade, business concerns costs
Technological Applied science, engineering	Ethical Values
Political Organizational or governmental	Personal Experience Personal observation, informal character

© 2001-- The Center for Teaching, Learning, Technology, General Education & The Writing Programs.
Washington State University

Appendix B. Rubric for Entomology 401 Term Paper: Spring 2000

Carol Sheppard, Department of Entomology, Washington State University (used with permission)

(Note that, except for #7, the bullets beneath each numbered item represent an incremental improvement in performance)

- (1) Identifies and explains *issue/topic at hand*
 - Does not ID nor explain main issue/topic at hand; is confused
 - IDs main issue, does not explain clearly
 - IDs main issue/topic clearly, explains in limited fashion
 - IDs main issue/topic clearly, explains fully by discussing subsidiary and/or other relevant issues
- (2) Identifies and uses a *primary, historical source*
 - Does not identify a primary, historical source, or cites an inappropriate one
 - Cites an appropriate primary, historical source, but merely repeats the information or does not engage it
 - Cites an appropriate source, presents/engages the information in a limited fashion
 - Cites an appropriate source, presents and engages the information, examines and assesses it
- (3) Identifies and considers *other salient perspectives/analyses* regarding issue/topic at hand
 - Does not cite nor utilize sufficient (or any) perspectives/analyses regarding the topic/issue
 - Cites and utilizes perspectives/analyses that are of limited value
 - Cites and utilizes salient perspectives/analyses, but does so in a limited fashion
 - Cites and utilizes salient perspectives/analyses, and brings them to bear on the issue/topic at hand
- (4) Identifies and presents the *student's own perspective/analysis* regarding the issue at hand
 - Fails to ID and state his/her own perspective/analysis on the issue/topic at hand
 - IDs and states own perspective/analysis, but fails to clarify own perspective versus other salient perspectives
 - IDs and states own perspective/analysis, but does so in a limited fashion
 - IDs and states own perspective/analysis, and considers it in light of other salient perspectives
- (5) Identifies and considers the *influence of context** on the issue/topic at hand
 - Does not present the issue/topic as having connections to other contexts
 - Presents the issue/topic largely within a single context (e.g., scientific)
 - Presents the issue/topic as having connections to other contexts, but in a limited fashion
 - Presents the issue/topic as having connections to other contexts important for the issue/topic at hand

- (6) Identifies *conclusions and implications* of the issue/topic at hand
- Fails to ID conclusions/implications of the issue/topic
 - IDs conclusions/implications, but within a single context
 - IDs conclusions/implications as having connections to other contexts, but in a limited fashion
 - IDs conclusions/implications relative to the contexts important to the issue/topic at hand
- (7) Follows “*Peer Review Guidelines*” regarding usage, composition, style, etc.
- Fails to follow established guidelines for usage, composition, style, and/or other requirements
 - Fails to provide list of references, or list is incomplete, or citations in text and reference list do not match
 - Fails to meet minimum page length required for term paper
 - Generally follows the guidelines listed in Entom 401 Coug Prints under “Peer Review Guidelines”

Contexts for consideration: scientific, technological, social/cultural, economic, political, ethical.

References

- Burke, S.M., & Kelly-Riley, D.O. (2002). *The Washington State University Writing Portfolio. Fourth findings: June 1999–May 2001* (Internal Report #5, January 2002). Pullman, WA: Washington State University, Office of Writing Assessment.
- Conrad, J. (1989, reissue). *Lord Jim*. New York: Penguin USA.
- Elbow, P. (1994). Writing assessment in the 21st century: A utopian view. In: L. Z. Bloom, D.A. Daiker, & E.M. White (Eds.), *Composition in the twenty-first century: Crisis and change*. Carbondale, Illinois: Southern Illinois University Press.
- Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication*, 28, 122–128.
- Facione, P.A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. ERIC Document Reproduction Service No. ED315423.
- Halpern, D. F. (1997). *Critical thinking across the curriculum: A brief edition of thought and knowledge*. Mahway, NJ: Lawrence Erlbaum Associates, Publishers.
- Haswell, R. H. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.
- Haswell, R. H. (1998a). Multiple inquiry in the validation of writing tests. *Assessing Writing*, 5 (1), 89–108.
- Haswell, R. H. (1998b). Rubrics, prototypes, and exemplars: Categorization and systems of writing placement. *Assessing writing*, 5 (2), 231–268.
- Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written communication*, 17 (3), 307–352.
- Haswell, R. H. (Ed.). (2001). *Beyond outcomes: Assessment and instruction within a university writing program*. Westport, CT: Ablex Publishing.
- Haswell, R. H., & Wyche, S. (1996). A two-tiered rating procedure for placement essays. In: T. W. Banta, J. P. Lund, K. E. Black, & F. W. Oblander (Eds.), *Assessment in practice: Putting principles to work on college campuses* (pp. 204–207). San Francisco: Jossey-Bass.

- Huot, B. (1996). Toward a new theory of writing assessment. *College Composition and Communication*, 47, 549–566.
- Jensen, A. (1994). Phlogiston, animal magnetism and intelligence. In: D. K. Detterman (Ed.), *Current topics in human intelligence*, Vol. 4: *Theories of intelligence*. Norwood, NJ: Ablex.
- Kohlberg, L. (1984). *The psychology of moral development: the nature and validity of moral stages*. San Francisco: Harper & Row.
- McLeod, S. H. (1992). Writing across the curriculum: An introduction. In: S. H. McLeod & M. Soven (Eds.), *Writing across the curriculum: A guide to developing programs* (pp. 1–11). Newbury Park: Sage.
- Murphy, S., & Ruth, L. (1993). Field testing writing prompts reconsidered. In: M. Williamson & B. Huot (Eds.), *Holistic scoring: New theoretical foundations, Advances in writing research* (Vol. IV). Norwood, NJ: Ablex Publishing Co.
- National Center for Education Statistics. (1993). *National assessment of college student learning: Getting started*. Washington, DC: OERI.
- National Center for Education Statistics. (1994). *National assessment of college student learning: Identification of the skill to be taught, learned and assessed*. Washington, DC: OERI.
- National Center for Education Statistics. (1995). *National assessment of college student learning: Identifying college graduates' essential skills in writing, speech and listening, and critical thinking*. Washington, DC: OERI.
- Paul, R. (1990). *Critical thinking: How to prepare students for a rapidly changing world*. Santa Rosa, CA: Foundation for critical thinking.
- Paul, R., Elder, L., & Bartell, T. (1997). *California teacher preparation for instruction in critical thinking: Research findings and policy recommendations*. Executive summary Study of 38 public universities and 28 private universities to determine faculty emphasis on critical thinking in instruction. Sacramento, CA: California Commission on Teacher Credentialing. Retrieved April 3, 2003, from <http://www.criticalthinking.org/schoolstudy/htm>.
- Smith, W. (1993). Assessing the reliability and adequacy of using holistic scoring of essays as a college composition placement technique. In: M. Williamson & B. Huot (Eds.), *Validating holistic scoring for writing assessment* (pp. 142–205). Cresskill, NJ: Hampton Press.
- Scriven, M., & Paul, R. (2003) *Defining critical thinking. A draft statement for the national council for excellence in critical thinking*. Retrieved April 3, 2003, from <http://www.criticalthinking.org/University/univclass/Defining/html>.
- Toulmin, S. E. (1958). *The uses of argument*. New York: Cambridge University Press.
- White, E.M. (1996). Power and agenda setting in writing assessment. In: E. M. White, W. D. Lutze, & S. Kamusikiri (Eds.), *Assessment of writing: Politics, policies, practices* (pp. 9–24). New York: The Modern Language Association.