This is to certify that the
dissertation entitled

Bridging the Gap between College and High School Writing in an Online Assessment Community

presented by

Merilee Griffin

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Higher, Adult, and Lifelong Education

Roger S. Baldwin
Major Professor's Signature

May 13, 2009
Date

BRIDGING THE GAP BETWEEN COLLEGE AND HIGH SCHOOL WRITING IN AN ONLINE ASSESSMENT COMMUNITY

By

Merilee Griffin

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree

DOCTOR OF PHILOSOPHY

Higher, Adult, and Lifelong Education

2009

ABSTRACT

BRIDGING THE GAP BETWEEN COLLEGE AND HIGH SCHOOL WRITING IN AN ONLINE ASSESSMENT COMMUNITY

By Merilee Griffin

An interactive Web site was developed and investigated to determine whether interaction on the site among college and high school teachers of writing could result in the formation of an interpretative community. An interpretative community is one in which meanings are stable (Fish, 1980), and which adjudicates student writing with a high degree of reliability. Interaction on this Web site produced movement toward reliable scores, increasingly shared perceptions of student writing, and development of common perspectives in talking and thinking about writing. Several recommendations for future uses of a such a Web site are provided.

This dissertation is dedicated to

Winifred Ruth Fisher Griffin

whose example was my beacon

ACKNOWLEDGEMENTS

Roger Baldwin, my major professor, steadily provided me with insight and intellectual challenges in his collegial, modest, supportive way that prodded my conscience and motivation to do well. His unquestioning acceptance of my commitment and ability kept me plodding through complexity, frustration, and boredom to the finish line. Dissertations are for tortoises, not rabbits, and Roger helped me morph from a rabbit-style start-and-changeup sprinter to the long-term duration jogger I needed to become.

My other committee members, Ann Austin, Nancy DeJoy, and James Fairweather, each gave me more time and thoughtful attention than most doctoral students could imagine, and they shaped my thinking and work and the eventual results of this project beyond the scope of my own capabilities.

Trudy Banta of IUPUI and Sharon Derry of Wisconsin both smoothed bureaucratic pathways so that I could take their courses as a Big Ten Traveling Scholar, and they provided me with the latest of theoretical and practical developments in their respective fields to round out the literature in the strange academic niche of my interest.

The HALE Cohort of 2005, large, diverse, and rollicking, formed a much-needed web of camaraderie and support. They also responded quickly to midnight cries for help with sources, editing, and moral support.

Most of all, my daughters, Carolyn and Leslie Beia, provided both tacit and expressed support when their mother decided to veer off in a strange adventure in midlife. The unquestioning acceptance of someone's need to reach to exceed her grasp is perhaps the greatest gift I have received, from all the persons named above.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The ability of the nation's high schools to prepare students for college-level work has been a continuing issue in education since at least 1983, when the report, *A Nation at Risk: The Imperative for Educational Reform,* was published by the National Commission on Excellence in Education. Chartered by U.S. Secretary of Education Terrence H. Bell, *A Nation at Risk* planted the phrase, "rising tide of mediocrity" in the public consciousness. One of the Commission's charges was to study the relationship between college admissions requirements and student achievement in high school. Among other findings, the Commission reported that:

- "Many 17-year-olds do not possess the "higher order" intellectual skills we should expect of them.
- "Business and military leaders complain that they are required to spend millions of dollars on costly remedial education and training programs in such basic skills as reading, writing, spelling, and computation.
- "More and more young people emerge from high school ready neither for college nor for work.
- "[W]here there should be a coherent continuum of learning, we have none, but instead an often incoherent, outdated patchwork quilt."

Nearly 25 years later, another commission, chartered by U.S. Secretary of Education Margaret Spellings, reported nearly the same results. *A Test of Leadership: Charting the Future of U.S. Higher Education,* criticized the nation's colleges and

universities for several shortcomings, but also pointed quite clearly to the inadequacy of preparation for college work by secondary schools as a strong contributing factor in the dismaying slippage of U.S. postsecondary education in the context of international rankings. The report concluded:

- "Among high school graduates who do make it on to postsecondary education, a troubling number waste time—and taxpayer dollars—mastering English and math skills that they should have learned in high school." (p. vii).
- "Inadequate high school preparation is compounded by poor alignment between high schools and colleges, which often creates an "expectations gap" between what colleges require and what high schools produce."(p. 7).

If anything, preparation in high school for college work seems to have deteriorated rather improved during the nearly quarter-century between the two reports.

Both *A Nation at Risk* and *A Test of Leadership* have been criticized for politicizing education and for advocating simplistic, ideologically-based solutions to complex problems. Nevertheless, both of them highlighted a fault of U.S. public education that has not been seriously refuted in the literature: the tens of thousands of locally-governed public school districts and the thousands of diverse institutions of higher education in this country generally do not communicate effectively with each other about expectations of student learning outcomes.

The data cited in *A Test of Leadership* came from a 2006 study by *The Chronicle of Higher Education* (Sanoff, 2006), which had commissioned a comparison of high school and college faculty of writing by Maguire Associates. The study found that "forty-four percent of [college] faculty members say students aren't well prepared for college-

level writing, in contrast to the 90 percent of high school teachers who think they are prepared."(p. B9). Only five percent of college teachers believe their institutions are very successful in making academic expectations clear to high-school teachers. The Maguire report concluded: "Over all, both teachers and faculty members agree on an urgent need for better communication and greater interaction among high schools and colleges." (p.B9) Even earlier reports pointed to a substantial gap between what college writing instructors expect their students to know and what high school teachers believe they should know (Condon, 1997).

Against this background, innovative ideas for bridging the gap between high school and college learning outcomes seem worth exploring, especially in the realm of higher cognitive skills, including writing, that are harder to describe and assess than content knowledge or lower-level skills. The research undertaken for this project focused on student writing, a subject of much complaint and criticism. It included the development and implementation of an innovative solution to the problem, as well as the formal study of its outcomes.

The research investigated the effectiveness of an interactive Web site designed to develop the understanding of 12^{th}-grade teachers of college-preparatory writing courses about expectations of writing ability that college writing instructors have of their first-year students. Standards of writing competency are notoriously difficult to describe in meaningful ways, partly because the process of teaching has become divorced from the process of assessment, and partly because assessments are inherently biased because they privilege some values over others (Huot, 2002). The very language of writing assessment, terms such as "logical," "convincing," and "articulate," are subjective and open to a range

of interpretations. The qualities of writing expected by college instructors are multiple, complex, and difficult to communicate through curriculum guides, rubrics, and descriptive statements, which focus on categories of analyses (Huot). Developing an understanding of the expectations for college writing requires a close examination of student writing samples, discussion with other professionals, engagement in the practice of evaluating student writing, and meaningful feedback from an expert group to the novice group. In other words, to become master teachers of writing, faculty need to participate in a community of writing assessment practitioners. The processes of this kind of faculty development are well established, but they are prohibitively expensive for most educational institutions (White, 1994). It may be possible to deliver such training to wide audiences in a cost-effective manner, however, through an interactive Web site. The research reported here attempted to learn whether the use of this particular kind of technology is effective for the purpose of developing a writing assessment community, and if so, to explain how the development occurred, and to discover some of the strategies and techniques that are most successful.

The study sought to answer two primary research questions. The first question was whether teachers who engage in collaborative assessment activities on a Web site built for the study would develop an "interpretative community" that allowed them to analyze and score student work in similar ways. The second question was how such a community developed, if at all, in the context of the online work. Definitions of interpretive community and collaborative work are developed in the literature review, which follows in the next chapter, and the research questions are more fully explained in the subsequent chapter on methods.

CHAPTER 2: LITERATURE REVIEW AND THEORETICAL FOUNDATIONS

The theoretical basis for this research lies at the intersection of four distinct academic traditions, which combined, point to the necessity of revolutionizing the way writing is tested. (Not incidentally, the combination of these traditions also points to the need to revolutionize the way writing is *taught* in most contexts, but that is a project for the future.) In addition, literature from the fields of writing assessment and collaborative online learning supplement the theoretical warrant for this study.

Four Traditions

All four of these traditions rest on a constructivist understanding of the nature of knowledge and knowing as developed by Michel Foulcault and others: knowledge does not exist as an objective entity apart from human comprehension; rather, knowledge is socially constructed by humans in specific cultural and historical contexts. Each period of history develops its own "conditions for truth," which determine what can be accepted as true (Foucault, 1970). Each of the four traditions is also consistent with Jacques Derrida's work on the deconstruction of language, in which texts are opened to alternative meanings by identifying dualisms that mutually define each other, such as subject/object, male/female, and good/evil. Meanings are unstable; language is "slippery;" and texts are always subject to various interpretations (Derrida, 1976). The four academic traditions described below all work from a postmodern, constructivist paradigm.

First, the concept of "interpretative communities" was formulated by Stanley Fish (1980), a contemporary literary critic, whose work is most closely associated with English departments on college campuses and social commentary in literary publications. Fish proposed that the nature and quality of texts can be judged with a high degree of accord by persons whose values and perspectives have been informed through communal discourses, and who thereby tend to interpret texts in similar ways. Fish's notion of interpretative communities acknowledges the inherent role of culture and history in producing common understandings of values, the criteria by which the meaning of a text is constructed and judged by the reader. Fish's work has been cited in several recent theoretical tracts in the field of writing assessment, including books on writing assessment by White (1994), Yancey and Weiser (1997), Huot (2002), Broad (2003), and Elliot (2003).

Fish's concept of interpretative communities overlaps to some degree with the literature on situated learning and communities of practice produced in a second academic tradition. Jean Lave and Etienne Wenger are educational researchers, but their landmark work evolved from ethnographic studies of apprenticeship akin to anthropological studies among tailors of West Africa, Yucatec midwives, naval quartermasters, meat cutters, and members of Alcoholics Anonymous. Their 1991 book on situated learning established the importance of participation in communities of practice in learning, and explained how "legitimate peripheral participation," in which newcomers gradually take on not only knowledgeable skills but identities as expert practitioners, is the foundation of learning. Wenger (1998) later expanded on the concept of communities of practice. Both Fish's concept of interpretative communities and Lave

and Wenger's concept of communities of practice emphasize the embeddedness of learning in social contexts that are laden with common understandings developed over time and shared among participants. These understandings are always changing as part of a complex dynamic among actors, activity, and relationships. Such social contexts are central, not peripheral, to learning.

Lave and Wenger's and Fish's theories as applied to writing assessment are supported by the work on validity emerging from yet another tradition, educational measurement. Lee Cronbach (1989) and Samuel Messick (1993), cognitive psychologists working in educational psychology departments in the latter half of the 20th Century, influenced the way test validity was understood. Separately, they responded to criticisms of cognitive testing that began emerging in the 1970's, as the writings by Michel Foulcault on the social construction of knowledge and by Jacques Derrida on the social interpretation of texts were transforming the social sciences and humanities. Both Cronbach and Messick sought ways to answer challenges to the validity of standardized tests, which were under attack for failing to account for the influence of culture and values. Their work eventually formed the framework for a new definition of validity that has been adopted jointly by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council for Measurement in Education (NCME) (1999): "Validity refers to the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests." The requirement that interpretation of scores be supported by theory refers to the many constructs currently in use in educational and cognitive science theories that view knowledge as socially constructed and that situate learning in social contexts.

A fourth academic tradition, collaborative learning, provides additional theoretical support for this study. Kenneth Bruffee's book, *Collaborative Learning: Higher Education, Interdependence, and the Authority of Knowledge,* was published in 1993 and revised in 1997. In many ways, Bruffee's thinking presaged the lines of thought that would emerge a decade or more later in the research on online learning communities. Building on the work of Thomas Kuhn, William Perry, and Mary Louise Pratt, Bruffee shows how collaborative learning is instrumental in *reacculturating* learners, allowing them to challenge and modify the assumptions and biases of the groups they already belong to, in order to adopt the language, values, and mores of a new group. Rather than relinquishing membership in previous groups, learners "nest" the communities they already belong to within larger communities. The process is essentially linguistic: learning the language of the new community is paramount to developing full membership in it. Reacculturation occurs most effectively in *transition groups,* where a small number of learners mutually explore and break down previously acquired biases and assumptions in order to adopt the larger perspective of the new community. As with interpretative communities and communities of practice, collaborative learning is based on a *nonfoundational* understanding of knowledge: knowledge exists and has authority within communities where it is socially constructed in an ongoing process. Bruffee's work is especially relevant to this proposal, since much of it was done in collaborative writing groups, and because of his emphasis on language. He agrees with the view of Vygotsky and Oakeshott that participation in conversation occurs *a priori,* and is followed by the ability to think, rather than the other way around. We participate in conversation first, and are able later to utilize the expressions of others in our reflections and synthesizing of

thoughts. Bruffee's description of how learning occurs in collaborative groups provides a conceptual foundation for the assessment activities in this project. Bruffee and Wenger (1997) agree that learning in social contexts is not merely the result of occasional or intentional practices, but rather is an inherent part of the process that occurs whether participants are deliberate about learning or even conscious of it. This kind of social learning that results from participation in collaborative learning in groups is a key element of the study.

Taken together, the concepts of interpretive communities, communities of practice, the new definition of validity, and collaborative teacher learning point to the necessity of situating assessment of student writing in social contexts and accounting for those contexts in the construction of assessments. All four concepts support a view of knowledge as socially constructed and of learning, especially language learning, not as a simple process of explication and transfer of knowledge from one individual to another who assimilates it, but as a social process embedded in specific cultural and historical contexts.

Collaborative Learning in Online Communities

A literature emerging from educational psychology and learning science departments concerns the theory of using the Internet to build and maintain learning communities. Although this field is a relative newcomer, it has developed a rich literature both in theory and in empirical research. This literature complements the theories of interpretive communities, communities of practice, validity theory, and collaborative learning in advancing the view that learning occurs in social contexts and that learning

must be supported through social interactions. Social processes must be a central focus of design when technology is used to extend educational endeavors, and cultural beliefs must be taken into account when designing or evaluating models of learning communities. Bielaszyk (2006), for example, lists cultural beliefs as the first of four dimensions (along with practices, socio-techno-spatial relations, and external influences) that form the Social Infrastructure Framework, a model for design of online learning communities. Nardi (2005) illuminates the internal processes of learning communities by applying Engestrom's Activity Theory, a complex, dynamic model that explains the interaction of individuals working together toward a common goal (*objeckt)*, often in spite of varying or even opposing individual motivations. The power and passion of *objeckt* is often neglected in management theory, but essential to understanding the processes operating in learning communities.

The importance of the social nature of learning is further developed by Akkerman et al. (2007) who distinguish between group cognition as defined by overlapping or similar individual knowledge (the cognitive perspective) and group cognition understood more as a kind of gestalt of shared knowledge, implicit values, and tacit understandings (the socio-cultural perspective). This emphasis on sociality as a fundamental element of learning is taken one step further by Paavola, Lipponen, and Hakkerainen (2006), who envision group cognition as something beyond either cognitive or social participation models. They describe a third metaphor, the knowledge-creation metaphor, in which group members build upon each others' contributions to develop new knowledge that could not have been constructed by single individuals working in isolation. The

knowledge creation metaphor maps closely onto the collaborative learning process described by Bruffee.

Three trends in the literature on collaborative learning in online communities are especially relevant to this study. First, several researchers caution designers to be aware that design can either foster or inhibit the building of community among participants, and that design privileges, either explicitly or implicitly, certain values over others. For example, online communities and learning games can be structured to enact such values as personal affirmation, social responsibility, and affirmation of diversity – or their opposites (Barab, 2007). A second trend in the thinking about online communities conceives of knowledge-building in such communities as more than the addition of the knowledge of individual participants (Akkermann, 2007; Paavola, Lipponen, & Hakkerainen, 2004). In typical learning groups, each member adds to the store of his/her own knowledge through the process of transfer as other members share their individual stores of knowledge. But knowledge communities often act to create new knowledge that goes beyond what individuals bring to the group. Often referred to as *group cognition* (Akkermann et al.) or *knowledge creation* (Paavola, Lipponen, & Hakkerainen), this process goes beyond previous constructs of learning as either assimilation (transfer of knowledge from one individual to another) or participation (learning through social practices) (Paavola, Lipponen, & Hakkerainen). This line of thought descends from the literature about the social construction of knowledge that includes Vygotsky's concept of the mind as essentially linguistic, Engestrom's model of activity theory, and Heidegger's critiques of Western thought (Stahl, 2006).

A third notion is that, as members of learning communities become familiar with the socio-techno environment and with each other, their interactions often shift from a vertical, top-down orientation to a more horizontal, democratizing orientation, where group members often bypass authorizing officials, appropriate the technology for their own uses, and adapt technical affordances for their own aims (Fischer & Giaccardi, 2004). In fact, the concept of meta-design seeks to design technology so that end users can become designers, modifying the software to their local and immediate needs.

Bruffee (1999) has suggested three additional issues for designers of software intended to promote collaborative learning. First, learning tasks should be directed toward achieving consensus. Second, programs should help participants learn the social conventions of working successfully in small, semi-autonomous groups. Third, in addition to information and tasks, programs should include assumptions and evaluative criteria that help people learn from each other. The design of a Web site, then, has important implications for social learning and the cultural values and assumptions that underlie it.

Theory Applied to Writing Assessment

The constructs of interpretative communities, communities of practice, validity theory, collaborative teacher learning, and online learning communities have an especially compelling implication for writing assessment. Although educational measurement specialists usually envision "good writing" as the embodiment of particularly desirable qualities that may be isolated and defined apart from any specific social, cultural, or historical contexts, the qualities of writing considered desirable at any

particular time by any particular group of people are in fact quite peculiar to specific contexts. For example, a comparison of essay scores on college admissions tests between the 1920's and the 1980's revealed quite different valuations of "knowledge of the canon" and "personal and authentic voice" (Faigley, 1997). Current methods of assessing writing in mass, standardized contexts aim to eliminate socially-constructed elements of writing such as prior knowledge, linguistic traditions, or values specific to a particular group, since these are viewed by educational measurement specialists as confounding variables. In doing so, however, they strip away the very social context that gives language its meaning, and they produce tests that are capable of discerning only the most superficial and limited aspects of student writing ability (Elliot, 2003; Huot, 2002). The practice of testing students on writing ability in this way, without regard for socially-mediated elements of language use, continues unabated by educational measurement specialists working in testing companies and state departments of education, in spite of the revised definition of validity adopted by the three most prominent professional associations, the AERA, APA, and NCME.

Ironically, the method of reliable scoring that made modern writing tests possible was developed as a solution to the validity problems of multiple-choice tests that dominated writing assessment from World War II to the early 1970's. These indirect measures of writing ability required no actual writing, but asked students to choose between correct and incorrect answers or "best" and alternate constructions of language. The tests were criticized by faculty as lacking validity because they did not require students to generate any writing, but merely respond to faults in writing prepared for the tests. The Educational Testing Service (ETS) developed holistic scoring of writing tests

with the intent of providing a more valid means of assessment. As originally conceived, holistic scoring did not use any kind of rubric or scoring guide under the theory, which grew out of the German notion of *gestalt*, that any piece of writing comprises an organic whole that is greater than the sum of its parts, and that writing is too complex and multi-faceted to be reduced to any single scheme (Elliot, 2003). The holistically-scored essays are still challenged by writing faculty of rhetoric, composition studies, and literary criticism, because assessments are conducted under rigorous testing conditions rather than socially-appropriate contexts, and scoring is done by minimally-trained college graduates with no required background in language, writing, rhetoric, or literature, and no established relationship with the student writers (Huot, 2002). The result is that such tests emphasize lower-level skills and superficial features of writing rather than the full panoply of skills, including critical thinking, reasoning, stylistic control, and rhetorical strategies, that comprise effective writing.

More recent theoretical work in the field of writing assessment, notably the work of Brian Huot (2002), has proposed that writing assessment be: 1) locally-controlled; 2) site-based; 3) faculty-led; 4) rhetorically-centered; and 5) supportive of writing pedagogy. These criteria are consistent with theoretical foundations developed by Fish on interpretive communities, by Lave and Wenger on communities of practice, by Messick and Cronbach on validity theory, by Bruffee on reacculturation in collaborative teacher learning, and by Stahl, Akkermann, Nardi, Bielaczyk, and others on online learning communities. Furthermore, a broad consensus appears to be emerging around Huot's framework among the most prominent scholars of writing assessment (Broad, 2003; Elliot, 2005); in fact, Norbert Elliot said of Huot's book, *Re-Articulating Writing*

Assessment for Teaching and Learning, that "exactly the right book has been written at exactly the right time" (293).

It should be noted that localized communities of practice, collaboration, or interpretation enscribe specific standards of writing for their own community rather than subscribing to externally-imposed standards developed by institutional or state authorities. The advantage of local control over the development of standards is that it fully accounts for the cultural and contextual factors that underlie valid and authentic assessment and provides a means for teachers of writing to consider the entire range of writing skills and perspectives that comprise good writing. Such assessments solve the persistent problem of validity. Furthermore, to the extent that the results of assessment are used to improve and guide pedagogy, such assessments support a broad, enriched curriculum and instructional strategies that help students develop control and skill in real contexts that present a full range of rhetorical requirements.

The downside of such locally-developed assessments is, of course, that each local community, however constituted, could potentially develop standards that are at least minimally different from all others, and it is likely that would happen. Major variations would probably emerge among institutions with widely varying missions and types of students as each institution developed standards most suitable for its mission and clientele, with the unfortunate result that interpretative communities become isolated from each other. Students from the same high school attending different institutions could be subject to quite different expectations, as would students who transfer from one school to another. (Of course, the same conditions are true today, with the addition of great variations existing from one classroom to another within many of the schools.) While

solving the problem of validity, therefore, locally-controlled, faculty-led assessments raise the problem of reliability: a sample of student writing could be analyzed and scored very differently by different communities. In fact, Kathleen BlakeYancey (in Yancey & Weiser, 1997) has described the history of writing assessment as a continual struggle between validity and reliability, with each criterion achieving ascendancy in different periods at the expense of the other. The problem of obtaining both highly valid assessments in local contexts and reliability across local contexts, might be solved by applying Bruffee's (1999) notion of reacculturation, in which learners do not abandon previously-held values and assumptions, but rather nest them within larger communities as local communities open their boundaries to each other. In this way, individual faculty members engaged in learning communities would gradually construct shared meanings and adopt the values of the larger group, and so form an interpretative community that is capable of producing reliable scores across institutions. As individuals are inducted into larger and larger groups and are joined together in an ever-widening circle of larger communities, learning continues and prior knowledge is nested within new knowledge.

Bruffee's thoughts about learning groups may also help explain why the earlier attempts to train teachers of English in holistic scoring were successful and often enthusiastically received. In the late 1970's, shortly after the development of holistic scoring by ETS, the National Council of Teachers of English (NCTE) embraced the method as an alternative to multiple-choice tests and promoted the learning and use of the method by classroom teachers. Early experiments with training secondary teachers of writing in scoring produced great enthusiasm for the degree of professional development that occurred almost invariably as a by-product of the training (Huot, 2002; White, 1997),

but the intensive face-to-face sessions, usually conducted over three full days, required resources of time and professional development budgets beyond the scope of most school districts, and they were abandoned. The unusually warm response among teachers who were trained in holistic scoring may possibly be explained by the fact that the training sessions immersed them in interpretative communities, communities of practice, or collaborative learning communities where they were able to participate in the creation of socially-constructed knowledge and develop their own learning. Many participants reported a greater understanding of writing and writing pedagogy as a result of the holistic scoring experience, and said they believed it improved their teaching. Empirical research on faculty development of college instructors and high school teachers as a result of training in holistic scoring has not been reported in the literature, but there is frequent anecdotal evidence that these training sessions often produced enthusiastic collegial dialogs that advanced individual learning and pleasantly surprised the participants themselves (Condon, 1997; White, 1994; Yancey & Weiser, 1997).

As envisioned in the late 1970's, holistic scoring would not only allow faculty to regain control of writing assessment; it would also improve the quality of teaching by arming faculty with sophisticated knowledge about writing (White, 1994). The NCTE publication, *The English Journal,* featured articles about holistic scoring in many issues of 1977 and 1978, but by 1980, articles on holistic scoring had nearly disappeared. The training required long hours of face-to-face meetings, usually over three full days, and it was therefore expensive and difficult to schedule. Edward M. White, perhaps the foremost leader of writing pedagogy and assessment on the national scene at the time, warned that the teaching profession must learn the methods of assessment and take

control of the process, or it would revert to the testing industry (Elliot, 2003). His warnings were not heeded, and the testing industry did assume control of writing assessment, including testing for local school districts, for state departments of education, for college admissions, and for professional schools. At the time of this writing, standardized tests of writing are used in all 50 states in K-12 education, and they are the primary reform mechanism of the education act known as No Child Left Behind. Furthermore, federal policy, the accreditation process, business interests, and the testing industry are all exerting pressure on institutions of higher education to adopt standardized measures of learning for undergraduate education. Today, over 300 colleges and universities subscribe to the Voluntary System of Assessment (VSA), an initiative of the National Association of State Universities and Land Grant Colleges (NASULGC). As members of the VSA, they are required to administer one of three approved standardized test to demonstrate learning of higher cognitive skills (College Profile, 2009). The number of members appears to be growing (Collegiate Learning Assessment, 2009).

Summary

The alternative to standardized testing of writing and other higher-level cognitive skills that was robust in the late 1970's engaged faculty in the same scoring practice of student work that is currently in use in the nation's major tests, such as the SAT, ACT, and GRE. It was generally well received by faculty, led to faculty development (at least according to anecdotal evidence), and was considered valid and useful for teaching (Elliot, 2003; White, 1994). The current literatures on holistic scoring, writing assessment, learning communities, communities of practice, interpretative communities,

and faculty development do not include any studies that cross the boundaries of all of these fields to investigate the possibility of integrating them in an effort to solve a fundamental and intractable problem in education: how to achieve valid, reliable, and cost-effective assessments of writing (and possibly other higher-level cognitive skills) that are eminently usable by faculty to improve the teaching practice. Specifically, no literature directly describes the use of online collaborative learning for the purpose of writing assessment, and assessment is described only indirectly as a means of developing faculty learning. Given the contributions to knowledge of these various fields, however, these possibilities are ripe for exploration.

Especially because standardized testing is so frequently viewed by policymakers, business interests, and others outside the academy as the primary means of bringing about rigor and alignment of curricula (Spellings, 2006), this may be an ideal time to test an alternative means of aligning high school and college curricula and injecting rigor into secondary writing instruction. A new method, which I have come to call "online assessment communities," is based on a new combination of established practices (learning communities, communities of professional practice, interpretive communities, teacher collaboration, validity, writing assessment, and online learning). It may serve as a preferable alternative to reliance on standardized testing. The following chapter on methods describes one attempt to create and study an online assessment community.

METHODS

This chapter is divided into several sections. First, the research questions are explained. The following three sections describe the sampling strategies, the Web site that was developed for the study, and the data collection methods used. The next section explains why a unique mixed-methods approach was necessary for the purposes of this study. This is followed by a section on data analysis, which provides a detailed explanation of the procedures used and presents a rationale for each. The chapter concludes by noting some important limitations of the study.

Research Questions

Two primary research questions formed the basis for this study, and two sets of supplementary questions were added for the benefit of future researchers and those who may wish to implement a similar project. They are explained in the sections below.

Primary Research Questions

Because the concepts of communities of practice, interpretative communities, collaborative teacher learning, and online learning communities overlap in important ways and, taken together, provide substantial theoretical support for this work, the primary research question was taken from a statement by Stanley Fish (1980), which captures the essential role of community as a determinant of perceptions, values, and beliefs:

> ...[M]embers of the same community will necessarily agree because they will see (and by seeing, make) everything in relation to that community's assumed purposes and goals: and conversely, members of different communities will disagree because from each of their respective positions the other "simply" cannot see what is obviously and inescapably there: This then, is the explanation for the stability of interpretation among different readers (they belong to the same community). (p. 15)

There were two primary research questions:

1) To what extent does an interactive Web site provide the amount and quality of collaborative learning necessary for college writing instructors and high school teachers of writing to form an interpretive community?

2) How has interpretive community, if any, formed?

Secondary Questions

Two sets of secondary questions were also addressed. The first set sought information about how this type of Web site should be designed to achieve the most positive outcomes:

- What volume of sample student papers is sufficient to produce enough learning for formation of interpretative community?
- What amount and kind of analysis from college instructors is sufficient for formation of interpretative community?
- How much dialogue between college and high school teachers is needed to produce enough learning for formation of interpretative community?

- How much practice in scoring do teachers require for formation of interpretative community?
- How much dialogue among high school teachers is sufficient for formation of interpretative community?

Finally, another set of questions addressed the value ascribed to the Web site by both the college instructors and the high school teachers. This set of questions is important to the study, because voluntary use of the Web site by both groups is essential to its eventual application in practice. If professionals using the Web site feel that the time and effort they devote to it are worthwhile in the light of what they have gained, it may have widespread applicability and usefulness. If users' assessment of the Web site is that is has minimal worth, it may be possible to encourage use of it only by offering attractive incentives. If very negative views of the Web site's value emerge, it is unlikely that it will have applicability in the real world of education, even with attractive incentives. This is a critical question. One result could be that high school teachers develop a much greater understanding of what college instructors expect of entering students, but if they judge the effort and time they spend on the Web site to be greater than the benefit derived, the project will have limited applicability. Incentives could be offered by state Departments of Education (credit toward continuing certification, for example) or by their own school districts (priority in assignment to desirable college-prep writing classes, for example). If the Web site experience is highly frustrating or meaningless, however, teachers will probably not participate in it regardless of the incentive.

Although the experience of college instructors in the project was not a direct subject of inquiry, it was important to describe their general attitudes about it. If every other part of the study "works," for example, but college instructors cannot be engaged to provide the student papers, commentary, analysis, and scoring that furnish the Web site, an interpretative community could not develop in a meaningful way. The questions for each group are:

- (For college instructors) Considering the amount of time and effort required for participation, what value is ascribed by them to the experience in terms of high school teachers learning of college expectations of student writing?
- (For high school teachers) Considering the amount of time and effort required by this study, what value is ascribed to the experience in terms of preparing students for college work?

Sampling Strategy

The study employed purposive sampling, both for the college and for the high school and college participants, partly because the samples were small (one college, four or five college instructors, and ten or fewer higher school teachers), and because there was no intention to generalize to a larger population (Miles & Huberman, 1994). The sampling was also criterion-based, since all the criteria described below were needed to provide a good trial of the method (Creswell, 1998).

The College

The study was centered at a small college that draws students largely from surrounding counties. An institution of this size is small enough that high school teachers

in the local region whose students often enroll there can be easily identified. Blue Water College is located in a small Midwest city. It is a community college in formal status, but functions as a comprehensive college in the region, since attendees can earn bachelor degrees in 14 programs, master degrees in 14 additional programs, and a doctorate in business administration, all offered by ten partner universities on the Blue Water College campus. Faculty at the college provided information about surrounding school districts that send a high proportion of their students to Blue Water College.

Larger institutions, such as Research I universities that enroll students from all over the state as well as out-of-state and internationally, were not considered for the study. Because high school students in the local area would be no more likely to attend that type of institution than other colleges or universities, the link between high school teachers and college instructors would probably be less immediate and more tenuous. Therefore, the motivation of high school teachers to develop shared expectations with instructors of a particular college might be lower. Because the focus of the study was on developing a shared sense of expectations of writing achievement among high school and college instructors of writing, a relatively high crossover between graduates produced in local high schools and freshmen enrolled in the selected college was desirable. It was assumed that teachers of 12th-grade students have at least some awareness of where their students matriculate. It was also assumed that writing instructors of first-year college students are at least minimally aware of where their students attended high school. If both groups of writing faculty were aware of a sizeable link between their current students and their students' previous or subsequent courses in writing, that awareness may be a motivating factor for participating in the study initially and for seeing the project to

completion. The choice of college was carefully considered with this factor in mind, and college personnel were interviewed to determine both where their students attended high school and how aware instructors of first-year students were of their origins. Both of these assumptions proved to be accurate. In answer to interview questions, all the high school teachers said a large number of their students attended Blue Water, and in four of those cases, BWC was listed first among the colleges named. Also, the four college teachers said many of their students came from the surrounding five counties, although the college attracts students from well beyond that region also. They generally agreed that a strong link in writing curriculum and instruction between area high schools and BWC would be beneficial for their students.

Another consideration in the selection of the college lay at the opposite end of the size scale: it was important that the selected institution not be too small. Small community colleges or liberal arts colleges, for example, are likely to have a small number of instructors, who could more easily communicate with each other about standards and develop shared expectations. It is also more likely that their disciplinary and professional backgrounds are similar and that they would bring similar perspectives to the teaching of writing. Some diversity among faculty was desirable because their commentary and analysis in the assessment of student writing should represent a rich range of perspectives. Without such diversity, the entire exercise is likely to be superfluous, without much applicability to actual teaching contexts. Therefore, another factor in the selection of a college for the study was that the faculty include a range of professional backgrounds and perspectives on the teaching of writing. Blue Water

College's Communications Department had 29 instructors of first-year composition, with quite varied backgrounds, at the time of the study.

The College Instructors

Four instructors of first-year writing were recruited to participate. The number four was not essential, but too few instructors would not have provided the range of perspective desired for the study, and too many would have created burdensome problems in scheduling meetings, and possibly in obtaining enough agreement about the scores assigned to samples of student writing. Being aware that each new member added to a group increases the number of interactions (and therefore potential conflicts) exponentially, feasibility of the study would be at risk by adding too many faculty members. Studies on small group dynamics have identified five members as the ideal number for this type of task (Bruffee, 1999), and would provide several perspectives and a range of opinion without bogging the study down in scheduling or philosophical conflicts. Only four instructors eventually agreed to participate in the study, however, so the sampling method in this case was opportunistic, taking advantage of opportunities present in the context (Creswell, 1998). Because their backgrounds and their perspectives and value systems about writing were sufficiently varied, that number worked quite well. In fact, because the recording work was accomplished in pairs that switched membership each week, four proved to be a good workable number.

The writing department chosen for this study was selected with care regarding the level of consensus college writing instructors have about student writing prior to the beginning of this study. A department that already had a well-established program of

blind review of student writing, for example, would offer little opportunity to explore the difficulties of agreeing upon expectations for incoming students, identifying a good range of sample papers, and producing analytic commentary on them, and would likely emphasize a limited perspective. On the other hand, a department with deeply divided philosophical and/or disciplinary approaches to writing could mire the project in unresolvable disputes. Therefore, an attempt was made to select a department whose faculty exhibited a range of perspectives, but was not so sharply divided as to be contentious. Informal conversations with all the instructors as part of the recruiting process revealed that the departmental culture was tolerant of diverse perspectives and fairly collegial. The department has a curriculum that guides instructors in developing course syllabi and writing assignments, but there is considerable latitude in the way it is applied, and there is no departmental review or grading process. The Chair of the Communications Department was helpful in describing the faculty as both collegial and diverse, and this proved to be accurate.

The four instructors recruited for the study had varied backgrounds. One had considerable formal training in rhetoric; another had extensive experience in teaching reading to both children and adults; and a third had been instrumental in developing writing curricula with high school teachers in the region. All four had master's degrees. Three had taught high school at some point and one taught full-time in a local high school as well as part-time at the college. They were all experienced, with a range of service from 14 years to 35. In spite of varied perspectives and somewhat different schemes for valuing the various features and qualities of writing, the four worked together in a generally collaborative and congenial manner.

The High School Teachers

Since the high school teachers did not have to meet face to face and did not necessarily have to come to agreements about the scoring of student writing, a larger group could be recruited. A group of seven to ten teachers was sought to provide a range of experience in terms of years teaching and degrees obtained, and perspective (emphasis placed on grammar, rhetoric, personal voice, etc.), but would not be too many to interview three times each. The teachers would ideally represent a range of teaching experience, academic preparation (BA, MA; and majors in English literature, composition studies, American studies, etc.), rural, suburban, and urban backgrounds, and perspectives on writing (emphasis on structure, voice, correctness, style, etc.). They were recruited from the greatest number of high schools possible, given that each school must have a record of sending a large number their graduates to the college selected for the study, and the teachers should be at least minimally aware that many of their students matriculate at the selected college. It happened that the intermediate school district surrounding Blue Water College consisted of five counties, with a total of 20 high schools, all of which graduate a large number of students who matriculate to Blue Water College. Recruiting the high school teachers, then, was simplified by the cooperation of the English/Language Arts specialist at the regional Intermediate School District, who already had a communications network with local superintendents, principals, and teachers of English. First, phone and email conversations with prospective participants, and later, the initial interviews, indicated that the teachers selected for the study would provide a good range of background and perspective. Seven teachers were recruited

originally, but one quit shortly after the beginning of the project, leaving six high school teachers who finished the study.

Each of the six taught at different high schools. One taught in a poor, rural district. Two taught in small cities served by a single high school. One taught in a fairly new charter school serving students who had had difficulty in public schools but aimed for college immediately following high school. Another taught in a school in a wealthy resort area surrounded by a poorer local population, so it had a higher tax base than most schools in the area, but a higher proportion of low-income students. All the teachers had master's degrees, but their years of service varied between one year and 29 years. Except for the teacher who had just finished her first year, all of them had taught at several schools, ranging from small rural schools to middle-sized cities. Two had taught in private as well as public schools, and one was currently teaching in a charter school. Their areas of interest and specialization also varied considerably. One had extensive training and lifelong involvement in theater. Another also taught Spanish and had traveled widely. One writes fiction on the side. Another had lived and taught in the Middle East. Two had served as journalism teachers and yearbook advisors. One was interested in student leadership development. Their perspectives on teaching writing were equally varied. One had far more background in grammar and style than the others and based much of his teaching on those aspects of composition. One was grounded in the six-traits approach to writing. Another had been influenced by the structure and rigor of a Catholic high school where she began her career. Two of them attended optional workshops and conferences frequently and placed a high value on being up-to-date on state curricula and standards. Another had also taught elementary and middle school

language arts and had a language-development perspective of teaching writing. Two of them had taught reading at some point. It seemed likely initially that the greatest difficulty in obtaining diversity of experience and perspective would arise from the likelihood that teachers assigned to 12th-grade college-prep classes would be among the most experienced and best educated members of the English department faculty and so have much in common, but the six teachers who agreed to participate in the study presented a good variety of backgrounds and points of view. Each teacher agreed to spend a minimum of 20 hours at the Web site, to be monitored during their initial hour online for the purpose of detecting and fixing technical problems, and to be interviewed for one hour immediately before beginning the project, about midway through, and at the conclusion of the project. They further agreed that the written record of their online discussion would be analyzed as part of the research.

Incentives for Participation. Two grants from the researcher's graduate school and department provided for small honorariums for the participants. The four college instructors each received $500, and the high school teachers received $200. In addition, the high school participants earned two State Board Continuing Education Units (SB-CEU's), which are required by the state's department of education to maintain certification.

The Web Site

Before the study could be implemented, it was necessary to design and construct the interactive Web site that would provide the working site for the study. Some

considerations about study's grounding in the constructivist paradigm must be explained before describing the design of the Web site.

Constructivism and Positivism in Design of Web Site

The theoretical framework for this study developed in the literature review clearly situates the project in the constructivist paradigm, where knowledge is viewed as a social construct rather than having an absolute existence apart from human comprehension. Ideally, then, any knowledge generated by the community of participants in the study would be formed without the imposition of any standards constituted *a priori* in some external reality. The logistics of this study, however, required a small departure from an ideal of dialogic interactions in which all participants have equal standing to contribute to knowledge developed through mutual engagement. Instead of creating equal footing for both the high school and college teachers, the design privileged the college faculty's knowledge and judgments about student writing above those of the high school teachers. This was necessitated by two conditions: (1) the college teachers were free to begin work several weeks before the high school teachers finished their academic year; and (2) the college teachers worked from a single location, whereas the high school teachers were distributed over five counties. The original conception of the project was to recruit equal numbers of teachers from each level and bring them together in a common dialog without any prior judgments about standards. The decision to bias the design in favor of the college teachers may be partially justified by the fact that students move from high school to college, rather than vice-versa, but it nevertheless runs counter to my belief that teachers at each level become knowledgeable about developmental needs and appropriate instruction for the students they teach, and rightfully should have equal access to

decision-making authority in the task of assessment. The final design, then, was not wholly a constructivist conceptualization of assessment, but instead privileged the judgments of the college teachers, whose analysis and commentary on the student papers were presented as "expert."

This element of design does not move the project wholly into a positivist paradigm, however. The privileged judgments of the college teachers evolved themselves from a dialogic process rather than being imposed from an external authority. Furthermore, the language of recruitment, orientation, and the interviews with high school teachers did not present the college teachers' knowledge as superior or absolute; instead, I used phrases such as "the way college teachers teach," "what college teachers value," and "how college teachers think about student writing." The college teachers themselves never injected their ideas as absolutes either; their language was invariably collegial and collaborative.

Nevertheless, the design clearly privileged the college teachers to a great extent, in contrast to instituting a process in which knowledge is accessible equally to all participants.

In this sense, the study might be conceptualized as existing along a continuum that runs from a purely constructivist approach at one end to a purely positivist approach at the other. The design, then, is much closer to a constructivist approach to knowledge and learning than to a positivist one, but must be understood as a compromise. I would very much like to study online assessment communities created in a purer form in the future, but a modified design was necessary for this study.

Design

The Web site was named LINC, for Learning In Networked Communities. It was hosted on ANGEL, the course management program used at the researcher's university.. Initially, the programming was done by a graduate assistant from the Center for Writing in a Digital Environment (WIDE) at MSU, but it proved simple enough that the researcher was able to complete it with telephone assistance from the MSU technical help desk.

LINC has three main parts. First, there is some orientation material that includes the philosophical and theoretical basis for the program and information about its intended uses. Second, there is a directory with links for the 25 student papers, the scoring table, the discussion forums, and the commentary on each paper by the college instructors. Third, an administrative area provided reports about each individual's participation. Screen shots of pages of the site can be found in Appendix A, p.163.

The Content

Twenty-five student papers were solicited from a variety of sources. All were actual samples of student work in first-year college composition courses, although a few were collected from other institutions as well as Blue Water College. The 25 papers were randomly assigned to five groups, or "Rounds" of five papers each, and each round was assigned to one of the five weeks of the study. The papers were numbered for both the Round and the paper within the Round: for example Paper 1.1 was the first, Paper 3.1 began Round Three, and the project ended with Paper 5.5. In each week, then, five student papers were examined, scored, and discussed.

First, the college instructors were asked to read each paper individually and jot notes about their analysis of it. The notes were then sent to the other three instructors, so that all four instructors had some input on every paper. Second, the instructors teamed up in pairs to prepare the college commentary and scoring. The pair would meet, review the notes by the other two instructors, and agree on a score for each paper and on the most relevant points for the commentary. The pair would then make an audio recording of their comments and analysis, ending with the score for the paper. Finally, the audio recording was sent to the researcher, who edited it and prepared PowerPoint slides that summarized main points and quoted parts of the paper to illustrate the points. Each of the papers had a slide show consisting of from 13 to 38 slides, with a mean of 21 slides per show. The embedded audio recordings ranged from a few seconds for the title slide to a few of over two minutes in length. Most of the recordings were in the range of 14 to 30 seconds. The audio portion began automatically, but the viewer could repeat the recording if desired by clicking on a speaker icon. All the slides used the same graphical template, with slight differences in background and font colors. Most of the slide shows began with a few introductory comments, and then developed specific points in detail by quoting material from the text of the student's paper. The score assigned to the paper was given at the end of the show.

The PowerPoints with embedded audio portions were then uploaded to the Web site. The four college instructors also agreed to monitor the discussion forums and respond with clarification and additional comments when appropriate, but they were asked to play a supportive role rather than taking a prominent role in the discussion. Pairings of the

college instructors were changed each week to prevent pairs developing divergent perspectives over the course of the project.

Working on LINC

At the beginning of each week, the high school teachers would log on, download and print the papers, score them, and conduct their own analysis, making notes to themselves. They would then log on again, enter the score for the paper on the scoreboard, and then join the asynchronous discussion forum to explain their comments, justify the score they assigned, read the comments of other participants, and respond to each other. Access to the scoreboard was controlled so that they had to make an entry before viewing the scores assigned by others. Likewise, access to the discussion forum was controlled so they had to enter a score before viewing the comments of other participants. Finally, after both scoring the paper and discussing it, they could access the PowerPoint slides to view and hear the commentary, analysis, and scoring of the college instructors for that paper. Five papers were thus treated each week for five weeks, for a total of 25 papers.

Originally it was decided to use a six-point holistic scoring scale in assessing the papers, because that is typical in college-admissions writing tests. Four-point scales are also common, however, and that was preferred by most of the high school participants since it is similar to the four-point scales they customarily use to grade papers. Switching to a six-point scale was confusing to some of the college instructors also, so the four-point scale was instituted shortly after the project began. The scores were immediately reported in the familiar language of college grading, however, so that half-points (1.5, 2.5, 3.5, etc) were assigned. Ranging from one paper that was given a 0.0 by the college

faculty to two papers that earned a 4.0, this effectively resulted in a nine-point scale (0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0). These finer gradations have the effect of making the desired interrater reliability of 8.0 more difficult to achieve, since scorers must approximate smaller intervals on the scale.

Strategic Decision

A decision had to be made before starting the project about whether to use a scoring rubric, which would provide a common framework and language and probably result in shared perceptions and expressions more quickly. That is the reason rubrics are used almost universally in scoring modern writing tests (Broad, 2003). A rubric would also introduce an element of authority into the process, however, and since it would have been prepared by the college faculty, it would enhance their already privileged position. Also, the primary research question was whether a shared "seeing" common to an interpretative community could develop in this environment, and a rubric almost certainly would have affected this process. While it may have led to greater common understandings, whatever conclusions could be drawn may have been an effect of the rubric instead of from the shared discourse. A rubric would also have introduced questions about its quality, its appropriateness for different pieces of writing, and how well the teachers were trained in its use (Broad), all of which could be confounding variables. Therefore, I decided it would be best to forego the rubric.

Data Collection

Ratings of Student Writing

A table was created from the online record of scores given by the college instructors and by each of the high school teachers. The table includes a row for each of the 25 papers and a column for each participant's scores, with another column for the combined score assigned to each paper by the college group. The scoring table appears in Appendix B, p.166.

Discussion Forums

All postings on the discussion forums were printed for analysis. Each paper is identified with the number of the Round, followed by a decimal and the number of the paper within the Round (Paper 1.2, Paper 4.2, Paper 5.3, etc.). Each comment is identified by the participant's pseudonym. With single spacing, the printout is just over 90 pages long.

Interviews

Each participating teacher was interviewed for approximately one hour immediately before beginning the project, about midway through, and at the conclusion of the project. A semi-structured interview protocol (Appendix C, p.167) for the first interview collected information about the teacher's background and focused on prior awareness of the expectations college instructors have for writing and on what features and qualities they thought were important in student writing. The interview protocol was developed in an iterative fashion, with additional questions for the later interviews being formed in response to data obtained in earlier interviews (Fontana & Frey, 2005). For example, if a particular concern was expressed by one or two of the high school teachers

interviewed previously, a question about that concern was added to the subsequent interviews.

Interview questions in the midterm of the project (Appendix D, p.170) focused on three kinds of responses: (1) how familiar the teachers were with the concepts of student writing expressed by the college instructors (or, alternatively, what surprised them); (2) how clear or how confusing they found the commentary and analysis of the college instructors; and (3) how useful or not useful they felt the experience would be for informing their own teaching practice with college-bound high school students. Since most of the content they read and heard on the Web site was developed by the college instructors, the interview questions had to be open-ended enough to capture whatever ideas were of interest to the high school teachers rather than structured around predetermined traits or qualities of student writing. As in the first round of interviews, additional questions could be developed during the course of interviews.

The final interview protocol (Appendix E, p.172) focused on summarizing the teachers' experiences and reflections on their experience in the project, especially on what kinds of benefits the experience provided them and what kinds of "costs" it entailed. Teachers were also asked to predict what influence, if any, their participation in the Web site might have on their future teaching of high school students. Again, after the prescribed course of questions, additional questions could be posed that were drawn from previous interviews with other teachers.

The four college instructors were also interviewed individually at the conclusion of the project (protocol in Appendix F, p.174). The questions focused on their experiences of discussing the student papers with each other, coming to agreement (or

not) on the salient features of the writing that should be commented on and taken into consideration when deciding on a score, the ease or difficulty of agreeing on scores, their impressions of the Web site, their reactions to the way high school teachers used the site, and their general evaluation of the usefulness and worth of this kind of communication. All the interviews were audio-recorded for accuracy, since all of the participants gave permission, and notes were taken. Table 1 on p. 37 pairs each research question with the source of data needed to answer it:

Table 1: Research Questions with Sources of Data

Primary Research Questions	Data Collection Method	Appendix & Question #	Analysis
To what extent does an interactive Web site provide the amount and quality of collaborative learning necessary for college writing instructors and 12th-grade teachers of writing to form an interpretive community?	Table of scores assigned to writing samples; discussion forums; interviews		Scores at the beginning and end of the study were compared to discern whether they became more consistent over time.
How has interpretive community, if any, formed, as evidenced by data in the interview transcripts and discussion boards?	All interview and discussion board data		Qualitative analysis; iterative coding of items to identifying categories and patterns
Secondary Research Questions			
What volume of student sample papers is sufficient to produce enough learning for interpretative community formation to occur?	Midterm interview; Final interview	App. D, #5; App. E, #2,3, 4	Movement toward consensus in the table of scores was compared with comments made at the points of greatest movement.
What amount of and kind of analysis from college instructors are sufficient for interpretive community formation to occur?	Midterm interview; Final interview; Discussion Board transcript	App. D, #6 – 9; App. E, #5 - 15	Movement toward consensus in the table of scores was compared with comments made at the points of greatest movement.
How much dialogue between college and high school teachers is needed to produce enough learning for an interpretative community formation to occur?	Midterm interview; Final interview; Discussion Board transcript	App. D, #6 – 9; App. E, #5 - 15	Movement toward consensus in the table of scores was compared with comments made at the points of greatest movement.

Table 1 Continued

Secondary Research Questions	Data Collection Method	Appendix & Question #	Analysis
How much practice in scoring do teachers require to form interpretative community?	Midterm interview; Final interview; Discussion Board transcript	App. D, #6 – 9; App. E, #5 - 15	Movement toward consensus in the table of scores was compared with comments made at the points of greatest movement.
How much dialogue among high school teachers is sufficient to achieve interpretative community?	Midterm interview; Final interview; Discussion Board transcript	App. D, #6 – 9; App. E, #5 - 15	Movement toward consensus in the table of scores was compared with comments made at the points of greatest movement.
(For college instructors) Considering the amount of time and effort required for participation, what value is ascribed by the college teachers to the experience in terms of high school teachers learning of college expectations of student writing?	Interviews at project completion	App. F, #6 & 7	The degree of valuing of this experience was ascertained; Results were modified or explained by material generated by other questions.
(For high school teachers) Considering the amount of time and effort required by this study, what value is ascribed to the experience in terms of preparing students for college work?	Final interviews	App. E, # 17 & 18	General impressions of value were summarized along with reasons given.
What features of this type of Web site appeal most and least to college instructors? To high school teachers?	Final interviews	App. E and App. F, all questions	Highly positive and highly negative comments were summarized.

A Mixed Methods Approach

The research questions and the data collection methods of this study lent themselves to neither a strictly quantitative nor a strictly qualitative methodology. This may be partly a function of writing assessment tradition, in which both quantitative measures such as holistic scoring and qualitative analysis are used (Elliot, 2003). It may also result partly from the opportunity to observe a phenomenon (formation of interpretive community) from both quantitative and qualitative perspectives.

Miles and Huberman (1994) present a cogent argument for the mixed use of both methods. "At bottom," they say, "we have to face the fact that numbers and words are both needed if we are to understand the world" (p. 49). The important question, Miles and Huberman write, "is not whether the two sorts of data and associated methods can be linked, but whether it should be done, how it will be done, and for what purposes" (p. 41). They suggest that two purposes for mixing of quantitative and qualitative methods are "(1) to offer corroboration and confirmation of each other via triangulation; and (2) to elaborate or develop analysis, providing richer detail (p. 41). Both of these purposes are served by a mixed-method approach to this study. The question of whether an interpretive community has formed can be investigated through the quantitative measures of holistic scoring assigned to the writing by the teachers, which is an established procedure in writing assessment. Richer detail can be provided, however, by a more qualitative investigation into the ways student writing was described and analyzed by the ten participants in the study. The two approaches are entirely complementary: one tracks movement toward interpretive community in numerical patterns, while the other describes the same movement in the perceptual frameworks and evaluative language of

the participants. Together they provide confirmation and elaboration of findings that would be lacking if either approach were used alone.

Analysis of the scores given to student papers by the participants was accomplished by tabulation and a statistic called intra-class correlation, which is explained in the section on procedures. Scoring was only one part of the study, however, and other methods were needed to analyze the discussion forum and interview material. Qualitative methods were appropriate here, but even so, none of Creswell's (1998) five traditions or other descriptions of qualitative methods in the literature exactly suited the needs of this study. The closest match was Creswell's explanation of grounded theory, which allows the researcher to develop categories of data using an established procedure of coding. Grounded theory also produces a theory of why or how a phenomenon occurs, however, whereas the goal of this study was a more limited inquiry into how people change in their perceptions and descriptions of specific attributes of writing and writing assessment as a result of community. Therefore, the procedures for grounded theory were used to a point, but the third step of developing an original theory was not applicable. Considering all the questions and the data available for analysis, it was necessary to develop an unique mixed-method approach that afforded the best interpretation of data in relation to the research questions. The procedures of analysis are described in the following section.

Analysis

This section explains the criteria and procedures used in interpreting the data. It is arranged in the order of the research questions.

Question 1: Indications of Formation of Interpretive Community

The question. The first research question asked was: To what extent does an interactive Web site provide the amount and quality of collaborative learning necessary for college writing instructors and high school teachers of writing to form an interpretive community?

The criteria. Two main criteria were used to determine whether an interpretative community has formed. The first was a quantitative matter, whether student writing was adjudicated by high school teachers with a high degree of reliability (or in Fish's word, stability) when compared with college instructors' ratings. Therefore, the scores given by high school teachers were compared to those of the college teachers for each paper in the study. If the difference between scores decreased over the course of teachers' use of the Web site so that they approximated the college instructor's ratings to a greater degree by the end, it would be taken as an indication that an interpretative community developed about the qualities of student writing (for example, strength of argument, organizational structure, and rhetorical strategies) that would be expected of "well-prepared" entering college freshmen.

A second indication of developing community is qualitative in nature, and could be inferred from the particular qualities and features of student writing that participants commented upon. If the high school teachers began noticing and pointing out features that they previously had not mentioned, but had been described and explained by the college teachers or other high school teachers, that too would indicate that participants were beginning to "see" things in similar ways. Instances in which the high school

teachers began commenting on student work in a tone or manner that more closely approximated the college teachers' approach would also indicate developing community.

The Procedures - Quantitative Analysis. A large body of research by both educational measurement specialists and researchers working in the tradition of writing assessment has been concerned with "interrater reliability," the degree to which trained raters arrive at the same or similar scores for student sample writings. Scorers typically work with four-point or six-point scales, with the highest number representing the highest level of achievement. On a six-point scale, then, a "six" paper demonstrates the highest quality of writing for a particular grade level or task, while a "one" paper demonstrates the lowest quality. In most scoring protocols, two raters read each paper and assign it a score without discussion or consultation with each other. If the paper receives two identical scores or two adjacent scores (for example, if one scorer assigns the paper a "five" and another assigns it a "four"), the results are considered reliable and the two scores are added together to produce the final score (in this case, a "nine"). If the two scores are more than one number apart on the scale, the score is considered unreliable, and it is adjudicated by a third scorer, usually someone with more experience and authority. Generally, researchers have established a correlation of 0.8 as the minimum threshold of interrater reliability, and they are generally pleased when the coefficient reaches 0.85 (Huot, 2002; Elliot, 2005; White, 1994). Although the interrater reliability correlations achieved in this study were noted in comparison with the "gold standard" of 0.8 to see how well reliability could be established by external standards, the main indication of community formation was whether a closer approximation of the scores occurred over time rather than the achievement of any particular coefficient.

For the quantitative analysis, the scores assigned to the 25 papers by the group of college teachers and by each high school participant were tabulated. The scores were converted from a grading-point scale, (1.5, 2.0., 2.5, etc.) to a nine-point scale (0.0=1; 4.0=9). For the purpose of being consistent in tabulating the scores, the following two rules were developed: 1) When occasionally a teacher reported a range of scores ("I'd give it a 3.5 unless this were a first-semester paper, and then I'd give it a 4.0"), the higher score was taken. 2) In two cases, the college teachers failed to arrive at a single score and reported two separate scores. Those scores were averaged when the result was a whole-number score on the nine-point scale, but when the result was a half-point ("6.5"), the higher number was taken.

When all the scores were tabulated, the scores given by each high school teacher were compared with the single score assigned to each paper by the college faculty to determine the interrater reliability for each teacher. Next, a mean of the high school teachers' scores was calculated, with the result reported to one decimal. The mean of the scores given by high school teachers was then compared with the scores given by the college faculty to determine interrater reliability for the group of high school teachers. The interrater reliability correlation in all cases was obtained by using the intra-class correlation method, which was appropriate for this study since two scores were being compared (individual high school with college; mean of high school with college); the data were continuous; and the sample size was small (Garson, 2008b).

The Procedures - Qualitative Analysis. The discussion forums and the PowerPoint slides for all the papers were printed for analysis. They were coded in an iterative process using the first two steps of the constant comparative method, open coding and axial

coding (Glaser and Strauss, 1967; Strauss and Corbin, 1998). (The third step, selective coding, was not attempted because its purpose is to develop a grounded theory explaining the data, while here the coding was used simply to detect any movement toward consensus from beginning to end of the project.) This was a difficult process, since over 75% of the material related to the qualities and features of successful (or failing) writing, and there are many ways to explain and categorize those features. For example, when a student mistakes "their" for "there," a teacher might describe the fault as a spelling problem, a grammar problem, a usage problem, or a homonym problem. Likewise, an abrupt change between paragraphs might be characterized by one teacher as faulty organization, while others describe it as a problem of transition language, as incoherence, or as faulty logic. Often, two or more teachers would comment on the same problem, but describe it in different ways. The various possible ways of categorizing problems meant that I could not simply rely on the language of the teacher comment to classify comments, but must make judgments about whether two comments referred to the same problem or to different ones. As noted in the subsequent section on limitations, this proved to be a challenging problem, and the findings must be interpreted with this difficulty in mind.

After the coding was complete, all the coded forum postings were tabulated. One table was constructed for each of the six high school teachers indicating which features the teacher had commented on in each paper. The fact that a comment was made about the feature was interpreted to mean that the teacher recognized the presence of it in the paper and believed it was important enough to describe (or that a feature was absent when it was considered desirable or necessary). When the tabulation of comments was

complete, each table was summarized to display totals for how many times the teacher had commented on each dimension and how many comments were made for each paper.

Comments on the PowerPoint slides containing the college faculty's analysis of the papers were also tabulated and summarized. Next, the six tables of high school comments and the table of college comments were combined to make comparisons easier. Finally, the combined Table of Comments (Table 9, Appendix I, p. 181) was examined to identify patterns, looking especially for features upon which high school teachers made no comments or sparse comments early in the project, but an increasing number of comments by the end of the project, and whether their comments more closely approximated the comments by college faculty over the course of the project.

After this process was complete, the discussion forum and PowerPoint printouts were read and coded in a second process. This second process also used the first two steps of the constant comparative method (open coding and axial coding), but this time examined data to discover another type of finding. This classification was not directed to features of student writing, but toward general perspectives of teachers on writing analysis and to interactions among the participants. Two of the perspectives that emerged from this coding process, for example, related to teachers' taking an instructional or helpful approach to comments that would aid student learning, and expressions of agreement or disagreement with others in the group.

The last procedure in the analysis for Question 1 (whether interpretative community developed) was to code the interview notes using the same categories that had emerged from the constant comparison coding procedures. The interviews were not transcribed completely, because the great majority of material was not relevant to the

research questions. I allowed the participants to talk about whatever they wanted to, so the interviews contained a lot of data that were not usable in the analysis. Instead, I located sections of the recordings that related to the research questions and transcribed those sections that related to the research questions. The interviews provided good background information about each teacher and some useful material that helped in interpretation of discussion forums, but they proved to be less informative than the discussion forums for answering the research questions. (A discussion of this phenomenon is included in the final chapter, Summary, Conclusions, and Implications.)

Question 2: Indications of How Community Formed

The Question. The second research question asked was: How has interpretive community, if any, formed?

Criteria. At various points in the project, participants' comments reflected a change of mind about how student papers should be assessed. Individual comments, exchanges among participants, and issues discussed just prior to or following these points of change might provide insight as to how or why the change occurred. Introspective comments, exchanges between two members about the value of features in student work, expressions of support or disagreement, and similar comments might indicate the motivation for changing opinions or formation of community, especially when those comments were followed by changes in the participants' comments, analysis, or scoring of papers.

The Procedures. In this procedure, I worked back and forth between the Table of Comments and the discussion forums. When the Table of Comments indicated there was a change in a teacher's pattern of comments at a specific point, I reviewed the related

sections of the discussion forum and the PowerPoint slides. I was looking for types of comment or interactions among participants made immediately prior to the change, or immediately afterward, that would explain how or why the change occurred. For example, when a teacher began commenting for the first time in a pedagogical manner, I tried to identify an exchange of opinions among participants, an introspective comment, or a specific comment from the college teachers' commentary that may have lead to the teacher becoming more attuned to pedagogical thinking. Interview responses were also reviewed for statements that might indicate why new or modified thinking occurred among participants. In the case of both discussion forum printouts and interview notes, analysis focused on the processes by which teachers came to a greater degree of understanding about how other participants perceived and valued student writing, and how their own scoring of student writing moved, if at all, toward consensus with others.

Next, the discussion forum printout was divided into separate sections, each of which displayed the comments by a single high school participant in sequence. These sections were printed again so that the comments made by each teacher could be read separately from the others. The Table of Comments was consulted to help identify patterns in each teacher's posts, and then a reading of each person's comments isolated from the others helped to understand each teacher's development over the course of the project. First, I read the material from the initial interview and the comments in the first three papers to develop a picture of the teacher's mindset toward writing assessment at the beginning of the project, with a specific focus on the features and qualities of student writing that were considered important. Next, I read through the remaining comments to the end, using the Table of Comments as a guide to identify points of change in the

teacher's pattern of comments. I chose quotations that illustrated the initial perspective, the points of change, and the final perspective, and wove them into a narrative that encapsulates the person's path of change from the beginning to the end of the project. Material from the interviews supported the interpretations of data in these narratives.

Question Set 3: Design of the Web site

A secondary set of questions addressed some design features of the Web site that may either facilitate or hinder formation of community. Although they are not as important as the two primary questions, the answers developed for them may aid future researchers.

The primary means of answering these questions was to review the findings obtained for the first two research questions, for the purpose of locating segments of time in the project at which change and learning among teachers continued and when it declined. If teachers continued learning throughout the project, it would be interpreted as an indication that the number of student papers and amount of discussion and commentary were appropriate. If learning declined at any point, it would indicate that the length of the project was too long and the requirements were too lengthy. (How much learning would have occurred if the project had been longer or used more material is a question for future research.)

In addition, the interviews with high school teachers conducted at the midpoint of the project and at the conclusion were read for information about the number of papers, kind of analysis from college instructors, amount of dialog between college and high school teachers, amount of practice, and amount of dialog with other teachers was most

conducive to formation of community. In cases where interview questions elicited closed-response sets such as "too many," "too few," or "about right," the responses were counted and reported. Most of the data about this set of questions was qualitative in nature, however, and was short and straightforward enough to be summarized.

Question Set 4: Participants' Attitudes toward the Project

The interviews conducted with both high school and college faculty at the conclusion of the study were read and classified to identify their attitudes, both favorable and unfavorable, toward the project, and to locate suggestions for improvement. I looked especially for any comments that would reflect a "cost/benefits" type of judgment about the value of professional learning received from the project compared with the amount of time and effort it required. These comments were tabulated separately for the high school participants and the college participants, and a few quotations typical of their responses were transcribed from the audio-recorded interviews for the purpose of illustration.

Limitations

The results of this study must be understood with several limitations in mind. Some of those pertain to the participants. First, it is possible that the teachers who volunteered for this project were exceptionally dedicated to improving their practice of teaching writing. Except for one person who had some very difficult family problems during the summer, all of them spent far more time on the project than required, and all exuded enthusiasm and professional commitment during the interviews. Also, as time went on, I began to suspect that the English/Language Arts teachers in this intermediate

school district had been exposed to more effective professional development than might be typical, since they were more knowledgeable about writing generally than recent reports on high school teaching of writing indicate (Chait & Venezia, 2009; Lumina, 2009; Spellings, 2006). It is possible, then, that the teachers recruited for the study were both more motivated and more knowledgeable than the general population of high school English teachers. As a result, the teachers may have started the study closer to the college teachers in attitudes and knowledge, and so had less distance to close on the scoring scale; or conversely, they may have moved a greater distance toward the college teachers than a more typical group of high school teachers would. It would be impossible to assess this effect without comparison studies with other groups of teachers.

It is also possible that the group of college faculty recruited for this project were exceptional. Although I had listened to their recorded commentaries many times over the summer in the course of editing them, dividing them into clips, and embedding them into PowerPoint slides, it was not until I was immersed in the coding process that I appreciated the high level of knowledge about writing and writing pedagogy they displayed. A less capable and energetic group could not have provided the quality of analysis and pedagogical approaches these teachers modeled, and the results would almost certainly have been less impressive. Also, the results of this study were obtained for a regional college with close ties to the surrounding secondary schools, but a more tenuous relationship that would be expected between a large university and a geographically diverse student population might result in quite different findings.

Another limitation applies to one of the participants. As noted, one teacher had exceptionally difficult circumstances during the study and did not log the twenty hours

online required to earn the two State Board of Education Continuing Education Units for certification. Although she finished the study, her responses were markedly abbreviated compared with others. Her comments do not demonstrate the same degree of change that is evident in the others, and the narrative developed about her reveals little of interest. For the purposes of explaining *how* community develops, then, only five participants are relevant to the analysis.

Another limitation concerns the student papers used in the project. It proved difficult to obtain both the papers and students' permission to use them during the summer, so there was little opportunity to select just the right range and type of papers from among a large group. We used whatever papers were available, and fortunately, they proved to represent a good range of ability typical of first-year composition students. The BWC first-year composition course emphasizes the argumentative/persuasive essay and the use of sources, a fairly conventional curriculum often called the "current traditional." The findings reported here, and conclusions drawn from the data, may not apply in programs with other emphases. In some cases in this study, we were not clear about the exact assignment or prompt that the essay was written to, and in about a third we did not have the Works Cited page, both of which affected teachers' certainty is assigning scores. Also, three of the papers were not of the same general argumentative/persuasive genre as the others, and that also affected analysis and scoring. All of these difficulties probably muddled the scoring process and may have clouded the findings. In addition to problems of assignment and genre is the confusion about how to categorize the various specific features of student writing. Although the Writing Program Administrators (WPA) have developed a nationally-accepted typology of standards for

college composition (Harrington, Rhodes, Fischer, & Malenczyk, 2005) they are not specific enough to guide the placement of detailed comments of the participants of this study into classifications for the purpose of coding responses. Therefore, readers should be mindful that the coding process was judicious rather than absolute.

An important limitation of the study is that I have a longstanding and deeply-held conviction about the teaching and assessing of writing: namely, that neither activity lends itself to an objective, positivist view of reality and learning, and that teaching and assessing student writing apart from social constructs is meaningless and even damaging. A common criticism of the use of standardized tests of writing in K-12 education, for example, is that the tests place an emphasis on superficial features of writing that are easily isolated from context, and so have the eventual effect of "dumbing down" the curriculum as teachers limit their instruction to those features that are detected by tests.

This bias could potentially influence the results of the study. One way I attempted to neutralize my participation is that during the period of the teachers' use of the Web site, I remained as invisible as possible. They had written notice that I was monitoring the discussion, of course, but I did not interject myself into the discussion at any time. My help was not called for in technical matters except for one brief episode at the beginning of the study. In fact, as events unfolded, I was so busy finding student papers for the project and creating the PowerPoint slides that I did not pay nearly as much attention to the discussion forums as I had planned. I did form relationships with several of the participants during the interviews, especially because many of their comments reflected my own experience teaching English and I was genuinely sympathetic with them. I expressed agreement and empathy frequently, but I never expressed disagreement with

their statements in spite of occasional opportunities to do so. In the data analysis process, I was guided by the numerical tabulation of comments to avoid identifying a change if it were not supported by the data. Nevertheless, it is quite possible that my biases and attitudes about assessment were apparent to the participants and influenced their interview responses.

Finally, the nature of this study does not fall neatly into a well-known genre of qualitative research, nor was it typical of quantitative research methods, so a mixed-method approach to data collection and analysis was developed to best address the research questions and find meaning in the data. While accepted research practices were followed where possible, a certain amount of adjustment and innovation was necessary, and the use of unconventional approaches to data analysis may reduce the trustworthiness of the findings.

FINDINGS

This chapter reports the findings from the several procedures outlined in the Methods chapter, organized by the sets of research questions in the order presented there. The first two questions were primary: (1) To what degree did an interpretative community form, and (2) if such a development occurred, how did the community form? The bulk of the analysis and findings below relate to those questions. The secondary questions addressed issues about the design of the Web site and how highly faculty valued it. The results for these secondary questions were comparatively simple and straightforward, and they are addressed at the end of this chapter. Some additional information about technical features of the Web site and preparation of the PowerPoint slides is included in Appendix L, p. 189, for future researchers.

Because so many quotations are used in this chapter, all of them have been separated from the rest of the text rather than following APA formatting, which would have mixed shorter quotations together with other text in paragraphs. I hope this improves readability. Occasionally, italics were added to quoted material to emphasize particular words or phrases, and those are noted as added to the original material. Also, the reader must know that the high school participants each chose pseudonyms for themselves to use during the study, and they are designated here by the initials FR, RR, DQ, SS, OR, and LL.

Question 1: To What Degree Did an Interpretative Community Form?

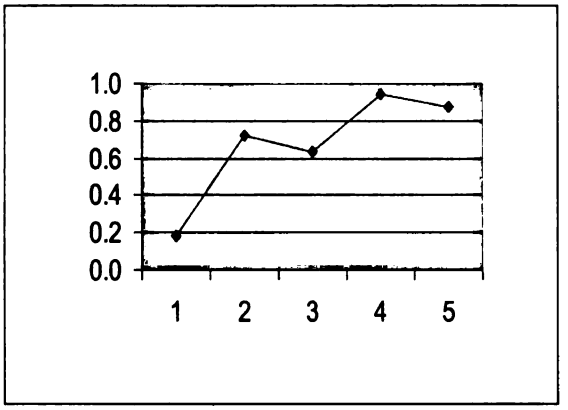
The data collected during this study point generally to a high degree of formation of an interpretative community among the teachers of writing engaged in the project. This finding is supported by both quantitative and qualitative measures.

Analysis of the Scoring Table

Evidence from the table of scores assigned to papers by high school teachers indicates a clear movement toward greater agreement with the college teachers' scores over the course of the study. After calculating the mean of the high school scores, the interrater reliability correlation between the mean high school scores and college scores was obtained for each Round by using the intra-class correlation method. The correlations are shown in Table 2, and graphed in Figure 1.

Table 2: Interrater Reliability

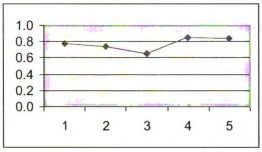
Round	Correlation
One	0.19
Two	0.72
Three	0.63
Four	0.94
Five	0.88



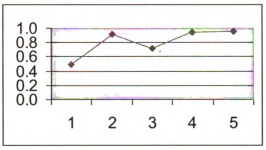
An intterrater reliability correlation of 0.8 is generally acceptable for writing tests in the testing industry, and a correlation of 0.85 or above is considered desirable (Elliot, 2003; Legg, 1998). The high school scores approximated the college scores with a correlation of 0.94 in Round Four, and a still desirable 0.88 in Round Five.

Similar results were obtained individually for each of the six high school teachers. Figures 2 – 7 show improvements in interrater reliability for scores of each of the six high school teachers over the five rounds. The underlying scores are in Appendix G, p. 176.

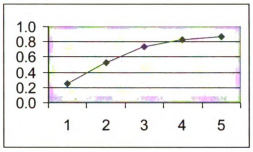
FR:



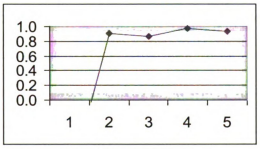
RR:



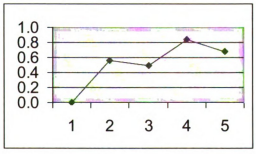
DQ:



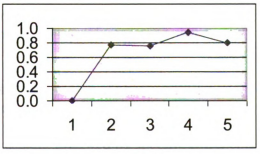
SS:



OR:



LL:



Two of the teachers ended the project with an interrater reliability correlation above 0.9; two were above 8.0; one was right at 8.0; and the sixth one below 8.0. All of the teachers' scores improved from Round One to Round Five, although the improvement was only slight for FR's scores, which started with the highest correlation. (For a note on the

negative correlation in Round One of SS's scores, see the underlying Table 8, in Appendix G, p. 176.)

Analysis of the Discussion Forums

Analysis using the constant comparative method of coding of the discussion forums printouts yielded similar results: clear indications of the formation of interpretative community emerged. In addition, some quantitative data compiled in the Table of Comments presented further supporting evidence.

Before reporting any findings from the discussion forum, a caveat is in order. An important consideration in the interpretation of this data is that, as time went on and participants became more familiar with each other, they frequently started a new post in the discussion forums by agreeing to comments made by earlier posters, rather than repeating what had already been said. These comments were coded as Expressions of Agreement. Five expressions of agreement appeared in Round One. These two comments were typical:

> I also agree that the inconsistent cite formats are distracting and the inclusion of very rudimentary ideas from dictionary.com are unnecessary and non-credible.
>
> I have to agree with FR that the Jetson's comment was awkward because it had very little to do with the rest of the paper.

Early in the study, agreement expressions were usually followed by specific comments that could be coded and entered into the Table of Comments.

As the study progressed, the number of agreement expressions grew to 18 in Round Four and 21 in Round Five, and more problematically, the teachers simply expressed agreement with previous comments and then went on to comment on other features rather

than repeating comments made in previous posts. Typical of Round Four and Five comments are these two expressions:

> You seem to have covered it all! I'm in agreement with everything you wrote.
>
> The mistakes are well enumerated already, and rightfully so.

This new style of agreement may indicate a high level of interpretative stability with each others' views and judgments that actually supports the idea the formation of interpretative community, but it frequently led to a dearth of comments that could be coded and entered into evidence, thereby weakening the degree of accord detectable in the quantitative record. The result is that the numerical tabulation of comments made about the various categories proved to be more useful as a guide to looking at the printed forum material, and as a check on my initial conjectures about results, either to support or to refute what I thought I was seeing, than as evidence itself.

I considered entering some value for expressions of agreement in the table of comments to indicate that another participant also found that "the topic was resonant" or "the thesis was clear," for example. Sometimes the expression of agreement was in response to an entire list or several paragraphs from different commentors, however, so the rationale for giving credit to expressions of agreement in the data became questionable. Therefore, the effort was dropped. If teachers had been asked to respond to a checklist or scoring rubric so that everyone had evaluated every dimension, the numerical data would most likely indicate much stronger formation of interpretative community than it does.

In the first step of analysis, the forum material was coded using the constant comparative method (Strauss & Corbin, 1998). In the open coding phase, attention was directed to the various features and qualities of writing that the teachers commented on. This part of the analysis was devoted to *features and qualities that appeared in student writing.* Eleven features emerged, as noted in Table 3. Throughout the rest of this report, the word "Features" is capitalized to indicate that this class of items is referenced.

Table 3: Features of Writing	
1	Topic
2	Introduction
3	Thesis
4	Argument
5	Support
6	Sources
7	Citations
8	Voice
9	Rhetoric
10	Conclusion
11	Conventions

After the Features were identified, two or more dimensions were identified for each Feature during the axial coding. The Dimensions (hereafter capitalized) were qualities of writing that could be arrayed on a continuum from a positive exemplar on one end to a negative exemplar on the opposite end. For example, the Feature of Voice exhibited three Dimensions: Consistent/Inconsistent; Authentic/Inauthentic; and Appropriate/Inappropriate. The Dimensions of each Feature appear in Table 4:

Table 4: Dimensions of Writing	
Feature	**Dimension**
Topic	Relevant/Resonant . . . . . Irrelevant/Not resonant
	Suitably narrowed. . . . . .Too broad or too narrow
Introduction	Engaging . . . . . Not engaging
	Apt . . . . . .Inappropriate
Thesis	Clear. . . . . Unclear
	Complex . . . . . .Simplistic
	Focused . . . . . .Unfocused
Argument	On point . . . . . .Off point
	Logical/Coherent . . . . . Illogical/Incoherent
	Explicit . . . . . .Vague or unstated
	Counter/concession . . . . . Limited perspective
	Depth/complexity . . . . . .Superficiality
Support	On point . . . . . .Off point
	Fully developed . . . . . .Undeveloped
Sources	Support claim . . . . . .Do not support claim
	Quality . . . . . Low quality
	Variety . . . . . Single source
	Integrated smoothly . . . . . .Not integrated
Citations	Correct . . . . . .Incorrect
	1-to-1 with references . . . . . .No 1-to-1 correspondence
Voice	Consistent . . . . . .Inconsistent
	Authentic . . . . . Inauthentic
	Appropriate . . . . . Inappropriate
Rhetoric	Devices/Strategies . . . . . No devices or strategies
	Arrangemt/Transitions . . . . . Weak arrangemt/Transitions
	Good diction . . . . . .Poor diction
	Sentence complexity . . . . . .Sentences short and choppy
Conclusion	Not just summary . . . . . Just a summary
	Relates to thesis . . . .Not related to thesis

Table 4 continued

Feature	Dimension
Conventions	Standard usage . . . . . .Nonstandard usage
	Spelling & homonyms . . . . . .Incorrect spelling
	Punctuation & capitalization. . . . Incorrect punct. & capitalization
	Grammar . . . . . .Incorrect grammar
	General . . . . . . Poor control or proofreading

A total of 35 Dimensions emerged from the coding process. Definitions for each one can be found in Appendix H, p. 177. After the Dimensions were identified and the printouts of discussions forums were coded, the codes were tabulated for each dimension, in each paper, for each high school participant. The individual tables were summarized to indicate the total number of comments made for each dimension throughout the project, and for each paper. The Table of Comments for each participant appears in Appendix I, p. 181. The codings of comments for the college group were also tabulated and are available in Appendix I. Some additional notes about how the discussion forums were coded can be found in Appendix J on p. 185.

The second coding process looked at comments that were not related directly to student writing, but reflect *perspectives or ways of talking and thinking about writing* that were exhibited by both groups of faculty. Findings about the Perspectives (also capitalized) will follow the next section on findings about the Dimensions.

Findings about the Dimensions of Writing

Of the 35 Dimensions of writing, three provided good evidence of increasing agreement between the two groups over the course of the project, all of them in the Feature of Argument. They included 1) Depth/Complexity, 2)

Counterargument/Concession, and 3) Explicit Connections. They will be discussed in that order in the following sections, accompanied by supporting quotations from the discussion forums.

1) *Argument Depth/Complexity.* The Dimension of student writing that the college faculty commented on most frequently was Argument Depth/Complexity, which was singled out for comment in 18 of the 25 papers. Depth/Complexity was defined as *"The writer does not stop at superficial, ideological, or cliched knowledge, but probes to gather additional information and presents an understanding of multiple perspectives."* Combined with the related dimension of Thesis Complexity, which was commented on in seven papers, these comments are evidence that an ability to deal with complexity in writing was an overriding concern of the college faculty. They gave much credit to a writer who recognized multiple or conflicting realities and was able to formulate a thesis that encompassed them, and likewise, they faulted writers whose understanding of the topic and presentation of the argument was simplistic. For example, they lauded the complex thesis in Paper 1.1 with this comment:

> The writer could have had a simple thesis that said, "Robots aren't going to be able to replace nurses." But the writer instead, in the second paragraph, said, "Fortunately, in the future I don't believe anything will really be able to take the place of the care that nurses give to their patients, despite how popular they are becoming, or how valuable their use is in other aspects of healthcare, like surgery." This more complex thesis gives direction to the paper and offers a forecast for what we can expect to read in the paper, and we think that's a really good way to write a thesis.

Conversely, Paper 3.5, which was a low point for complexity in the study, elicited this comment from the college faculty:

> The thesis itself, that celebrities make unacceptable role models, is not argumentative, when you're past the age of maybe 14, so it's not a thesis that we would find acceptable for a college paper.

The following commentary on Paper 1.3, on government regulation of the auto industry, illustrates how detailed the commentary of the college faculty was in dissecting the student's arguments and analyzing their strengths and weaknesses. It also shows how the faculty engaged directly in assessing the specific evidence of the argument.

> The writer does, however, try to make a case that there is some evidence that electronic stability control in new cars may be ineffective as a safety feature, and it's too expensive. The evidence for this expense is cited briefly as $250, which, when taken with the cost of a new car, is not such a great expense, thus weakening the overall argument.

Fourteen of the 38 slides on this paper addressed faults with the writer's handling of the complexity of the topic at a similar level of specificity. Throughout their analysis of the 25 papers, the college faculty consistently scrutinized arguments and theses at this level of detail.

An important shift occurred over the course of the study in the way the high school teachers treated this Dimension. Following are some quotations from the discussion forums that contrast each high school teacher's post on a Round One paper with a post from Round Five. For the sake of clarity, only the comments that pertain to thesis, argument, and related issues are quoted here. Comments about other Dimensions such as style and punctuation have been deleted from these quotations so the reader can focus on the Dimension of Complexity. Each participant's pair of quotations appears below, with the Round One comment quoted first, and the Round Five comment following it. Three asterisks separate one participant's pair of quotations from the others. It is not necessary to read all the quoted material to get a sense of how extensive and detailed the high school teachers' analyses became by the end of the study.

1.2

I liked the subject matter very much. The scenario introduction could have been much more effective . . . there is a thesis in there somewhere, but I had to read and then re-read to realize what it was.

The teacher questions the strength of the thesis and how it plays out in a subsequent paragraph in the first part of her analysis of a paper at the end of the study.

5.4

I felt that the thesis could have been strengthened also. Writer wasn't sure if he/she was focusing on security blanket or instant gratification. Writer continues his/her initial muddling by inclusion of a thought about creating a disregard for common manners.

Next, she questions the appropriateness of specific supports in light of her own experience with teen car problems, questions a transitional phrase, and questions the relationship of a paragraph to the thesis:

Para. 2—Idea about safety net coincides with my own thinking, so I found it particularly gratifying. However, the jump to rape, drug abuse, and wild partying was too big for me. When I think cell phones and safety nets, I think "flat tire" or "out of gas." I thought it valid that the author questions the strength of that safety net, but was put off by the almost-condescending tone in this sentence: "Let us be practical and assume the reality of teen behavior." Most importantly, this information seems out of line with a thesis which tries to focus on instant gratification.

Also, what is the purpose of the introductory phrase "in a fictional sense"? The writer does try, however, to get back to "instant gratification."

In addition to checking statements against the thesis statement, the teacher is weighing various statements against her sense of what is real and what is important in these posts:

> Para. 3—This paragraph begins with fluff about being graced with modern technology. I felt that the following ideas were much stronger (cell phones impairing resourceful thinking), but again misses the "instant gratification" mark.
>
> Para. 4—Overuse of cell phones can be fatal: quite an overstatement, as written, and the support about car accidents being the #1 cause of death doesn't quite match the assertion that cell phones cause accidents. Author remembers (almost as an afterthought) to include ideas about "the pursuit of instant gratification" and fulfilling "instant gratification."

The teacher continues weighing statements against the thesis and notices that, in spite of likable thoughts, the writer has veered off course:

> Para. 5—I like the sentiment behind cell phones crippling teens' social skills, but this paragraph veers entirely away from instant gratification. Writer almost begins a new topic here (patience is a value of the past), but I do understand the connection to his/her instant gratification thesis. The last sentence about forgetting values that make them human seems an odd, ill-fitting thought, as the writer is now focusing on ALL humans, not just teens.

In the final post, the teacher notices that the writer has again veered off topic, but compliments the writer on some general observations.

> Para. 6—Writer continues focusing on instant gratification, but now also has expanded into new subtopics (trading off values, responsibility, accountability). Observation about discontinuation of pay phones is a valid one, but off-topic. In spite of what I consider a disorganized paper, I like much of this writer's introspection ("society has no time for those who do not keep up"), especially the final thought about humans being reduced to little more than the sum total of their own flawed technology.

* * *

In his initial post on the first paper of the project, this teacher makes general statements about the title, opening, and thesis, but nothing else:

1.1

> I liked the attention-getting title and opening; I would like to see the thesis appear in the opening paragraph; a thesis does emerge and the conclusion works for me.

Toward the end of the project, this teacher observes the same features, but then begins questioning the writer's intent and stance. He then questions the use of the writer's terms, distinguishing between two close meanings:

5.4

> The title, introductory paragraph, and conclusion all coincide. I think I have a clear vision of this author's purpose; I am less sure whether or not the author approves of the new trend of teenage pregnancy. Merely pointing something out does not satisfy my notion of a controlling idea, of some judgment the author has made that justifies writing the paper in the first place.
>
> I notice in the second paragraph a need to make the distinction between *typical* and "normal." By using "normal" instead of typical, the author tacitly gives credence to teens' having babies. I am not sure at this point if this is the author's intent. I am not sure if the author knows she is writing about typical behavior that is not normal, but I suspect she does not know.

In the next post, the teacher observes his own response to a badly-phrased sentence, considers which meanings are clear and which are ambiguous, and then checks his own reactions to the statement. He also continues weighing assertions against evidence, and recognizes his own inference:

> The sentence "When one has a baby it is thought of to be bad but when a lot of teens are having babies, it becomes normal' is an abomination, but I do get the point, so communication occurs. That the taboo against having babies has been lifted is clear enough; nearly all the accompanying fears, shames, and excommunication from polite society have been deflated. I am still not sure whether the author approves of this or not. I appreciate the author's attempt to justify her position with evidence (outside sources). The author's attempt to imbue voice by means of relating her personal experience is the closest she has come in revealing her judgment about teen pregnancies—I infer she is mostly

> supportive, but perhaps she feels compelled to agree to lend an overall credence to her argument.

In this section the teacher evaluates the strength of the argument, deliberates about the writer's intent, and again, is aware of his own attempts to interpret the writer's words. He also evaluates the counterargument and the evidence of personal experience:

> The fact that she overstates her argument on page two continues my belief she is supportive: "Nobody is waiting until...." The rest of the paragraph supports teen pregnancy. The second paragraph on page two attempts to show the other side: that the movies really are heavily one-sided; again, the author uses her voice of experience—her best friend saga—to validate herself as someone who knows of what she speaks. However, this effort is feeble, since it ends with "happy grandparents."

In the next two paragraphs, the teacher evaluates a claim and the writer's failure to interpret it, along with a comment about the lack of sophistication about integrating quoted material into text:

> The outcome of this paragraph is that parents would rather prevent teen pregnancies and are willing to make certain concessions; the author does not interpret this information so I am not clear what she thinks. I got the impression this paragraph was thrown in to meet the research paper aspect of the assignment. Students often get lost in their zeal to satisfy what they view as research paper requirements.
>
> We get more of the same in the next paragraph on page three, finishing with the lifeless, irrelevant "Kristin A. Moore and Barbara W. Sugarland are writers for a journal."

At this point, the teacher is thinking not just about the argument, but about the motivations of the writer, who he thinks may be more concerned with fulfilling the assignment than making a point.

> Page four continues research that is not clearly connected. (Perhaps the author feels she is really just fulfilling the research aspect of the assignment right now). The trite remark "This could happen to anyone" and overstatement "Having a baby doesn't really matter..." uses persuasion, not convincing, to support her point—teen pregnancy are on

> the rise—a point perhaps she feels compelled to agree with in order for her paper to wash?
>
> The paper ends with a hearty endorsement that "pregnancy is now cool and exciting at any age." At least I finally get my long-awaited judgment.

The teacher now takes a longer view of the essay, identifying a faulty premise that weakens the entire work:

> My biggest complaint is that the paper is built around a faulty premise: the support for teens can end at any time; hence, one cannot predict the trend will continue. To suggest this may even be a positive social trend is naïve; in fact, teen pregnancies presently are having devastating consequences for many, in spite of the present support systems. The other side just doesn't get a fair shake.

The teacher who posted these comments not only analyzed some specific evidence of the argument in great detail, but took a long view of the piece in evaluating the premise.

* * *

Another teacher made these comments about the second paper in the project. She makes valid points about the argument, but at a high level of generalization, without any specific references to text:

> 1.2
>
> The writer's voice is fresh and authentic, and her ideas are pretty strong...The writer also points out that the men are rude to one another, which sets up a good point about this not being quality time, but then never follows that idea through.

This post by the same teacher on a Round Five paper contains references to more specific elements of the essay:

> 5.4
>
> I think I am in the minority here about the actual subject. I found the writer to be offensively simplistic and to overgeneralize a great deal. In paragraph 4 (page 3), the writer uses "(i)t is a well known fact" in two

sentences. In both of these cases, the writer then connects the "well known fact" to something that does NOT logically follow.

In this paragraph, the teacher weighs the writer's assertion against her own experience, and then tests the validity of the writer's argument:

> Many of you noted that the "instant gratification" idea wasn't very closely followed throughout the paper. I found particularly offensive the idea that cell phones will limit today's teenagers' problem-solving abilities. I have to admit that when I was 16 (eons ago) that I solved my car problem the same way today's "instant gratification" teenagers do – except I had to find a pay phone to call my parents for help. I don't think that the lack of a cell phone gave me better problem-solving abilities. I have managed to somehow grow up and become a competent problem solver, and I'm fairly certain that my sons will someday be competent problem solvers as well. Actually, they are sometimes more creative and able than I was at their ages. I am not saying that I like cell phones, and I see many problems generated by them, but I don't think that there is a valid connection between having a cell phone and being a poor problem-solver. (I think that teens in general, then and now, are just poor problem-solvers – probably due to lack of life experiences!)

In comparison with the first example, the teacher who posted these comments had entered the student writer's world of discourse. She is considering the merits of specific points of the argument in terms of her own logic and experience.

* * *

As in the previous example, this teacher's posts on the second paper made valid points, but did not pay attention to any specifics of the argument:

> 1.2
>
> It did have me pretty interested from the start. There was an attempt to grab the reader's attention . . .However, it quickly lost momentum, and it seemed as though the [writer] ran out of time and subject matter. Many generalizations were made later in the paper that lacked support ... Finally, the fact that there was no conclusion paragraph really made it fall flat.

The contrast between the previous example and the Round Five post is not great, but the later post evidences some consideration of points that do not appear in the Round One analysis:

> 5.3
>
> For me, the paper had the personality that the "Redesigning Recycling" paper was lacking, but it didn't have as strong of an argument. I'm not sure that it is possible to have the answers to such a big problem; however, the idea that the pharmaceutical companies should take back unused medications fell a little short of persuasion. What would be the incentive? Would this really be possible? Perhaps it is not the writer's responsibility to answer these questions, but it made the argument seem more hypothetical than realistic. Despite this, the paper was well-researched, well-written, and fulfilled the expectations of the assignment.

Although the teacher does not sustain a point-by-point examination of the argument in this Round Five post, she at least mentions some specifics and tests the student's premise against her sense of reality. She begins by applauding the paper's "personality," then thinks about the plausibility of the argument.

* * *

In this sample from postings on the first paper of the project, the teacher makes specific comments only about the thesis and seems to give up when immediately-identifiable topic sentences do not appear as signposts to guide the reader:

> 1.1
>
> I agree with RR that the thesis finally appears as the second to last sentence of the paper, though the argument was hinted at in the second paragraph. The writer seriously lacks topic sentences that would further his/her claim about robots serving as nurse assistants rather than replacements.

The same teacher begins her last post of the project thinking about the student. She has obviously read some earlier posts and is disagreeing with some earlier statements by

other teachers that criticized the paper. She goes on to consider the fairness of the student's argument and the rightness of taking an unconventional stand:

> 5.5
>
> From reading what the author had to say, we realize that he is a new college student with a lot to learn about mechanics, grammar, and proper essay writing (maybe he didn't get such an excellent education in the old building! Just kidding). However, I did not grade him harshly on these flaws because I think he makes a lot of really good points in his discussion of a real issue that our state faces: outdated school buildings. It's okay with me that he begins his paper seeing both sides of the issue and then later comes to take a solid stance. I do believe that this organizational structure works in his case because he wants to gain reader sympathy by staying neutral throughout examining the issue at hand.
>
> Furthermore, it is an interesting question: How does the quality of the building (and lack of modern technology) affect student learning? I think the author fairly considers both sides...I don't think he should be punished for his non-traditional approach to taking a stance later in the paper. Just because it doesn't fit our accepted model, doesn't mean it is wrong or ineffective.

Although this post is fairly short, there is considerable evidence that the teacher is considering a number of points (the student's current position and needs, the importance of the topic, the rhetorical strategy of delaying the stance, the unconventional approach) that fall outside the limited view of the comments on the earlier paper.

* * *

This last example contains pairs of statements from the beginning and end of the project that are not noticeably different from each other. Both of them make high-level generalizations about the papers, without offering any specific text as supporting examples, and neither one considers more than rudimentary elements of argumentative writing:

1.2

> The reading was a good topic. It needed much more development. .It clearly used opinions but was not using any facts or citing any sources...

5.4

> This paper is lacking any support for the arguments. The arguments were extremely weak and simply comes off as unfounded opinion.

In this case, written by the teacher who spent little time on the project compared with others, the comments show no evidence that the teacher's perspective at the end of the project was different from her perspective at the beginning.

* * *

Without even reading the text in these posts, it is evident that all the teachers, except for the last example from the teacher whose participation was limited, responded to the student's argument in far more detail in Round Five than at the beginning of the project. A closer reading reveals another difference: all the posts from Round One are summary judgments, as if they had been made in response to an inventory checklist, and a short list at that. There are no attempts to justify or explain the judgments in terms of the issues the student is addressing. For example, "I liked the subject very much;" "I would like to see the thesis in the first sentence;" and "Her ideas are pretty strong," express judgments, but not rationales for them. In the Round Five comments, teachers are fully engaged with the subject and are dissecting the arguments, as in these examples from the quoted material above:

> Overuse of cell phones can be fatal: quite an overstatement, as written, and the support about car accidents being the #1 cause of death doesn't quite match the assertion that cell phones cause accidents.

I notice in the second paragraph a need to make the distinction between *typical* and "normal." By using "normal" instead of typical, the author tacitly gives credence to teens' having babies.

How does the quality of the building (and lack of modern technology) affect student learning? I think the author fairly considers both sides.

In the Round Five examples, teachers are no longer making summary judgments about writing, but have actively entered the world of the argument. They are thinking through the logic point by point, questioning the strength of the evidence, and identifying the particular statements that do or do not hold up to scrutiny. It is as if they have entered a dialog with the student writer about the issues and are insisting that the student live up to high standards of argumentative discourse. The difference between the early and late responses of five of the six teachers could be characterized as a difference in perspective: in Round One, teachers are outside the universe of the student's discourse, observing it objectively and evaluating it. In Round Five, they have entered the student's world of thought and are engaged in a kind of jousting with the argument themselves. Such a perspective mirrors the approach taken consistently by the college teachers. It is probably also the approach that will be most helpful in explaining to students what is wrong with their overgeneralizations, leaps of logic, failure to connect dots, and so on, and thereby to help them become more critical about their own language and logic.

2) Counterargument and concession. This Dimension was defined as, *"The writer acknowledges weaknesses in his/her argument, incomplete or contradictory evidence, valid points made by opponents, and multiple viewpoints. Points are conceded and conclusions are qualified when appropriate."* The second-semester composition course at BWC includes a requirement to address counterarguments to one's own thesis and acknowledge valid points in opposing viewpoints. Students are required to research their topics adequately

so they are familiar with opposition arguments and to incorporate responses to them in their own papers. They are also encouraged to acknowledge evidence and reasonable arguments contrary to their own position when possible. In the tradition of Carl Rogers, for whom Rogerian rhetoric was named (Coe, 1992), counterargument and concession are believed by its proponents to strengthen arguments through reasonableness of dialog. This Dimension was introduced in the college faculty's analysis of the very first paper, on the use of robots in nursing, in this comment:

> We think there is some opposition. The writer does address the opposition, by making a concession, and acknowledges in paragraph 10 that robots are going to be in surgery in the future, and that maybe much of the surgical procedure will be carried on by the robots themselves, so that does offer a counterargument that robots *can* do some work that nurses currently do, and it's a nicely-handled concession.

The college commentaries mentioned counterargument and concession in all five of the Round One papers and in half of the remaining papers, often commenting upon their presence or absence and strength or weakness in detail. By contrast, none of the high school teachers commented on it in the first round, except for FR, who acknowledged the value the counterargument in the third paper in Round One, which she described as "weak." Rogerian argument was probably a new idea to the high school teachers: it was not mentioned in any of the initial interviews, and it is not generally featured in high school writing curricula. In fact, it is probable that comments made by two of the teachers in paper 1.1 (posted prior to accessing the college commentary) mistakenly interpreted a concession as confusion on the part of the writer:

> Furthermore, he/she seems unable to take a firm single stand on the value of robots in the health care profession.
>
> I also felt that the writer was wishy-washy about his/her stance on the topic. While much of the paper was trying to prove that human nurses

> were better than robots, there were places that it seemed that the opposite was being stated.

In Round Two, there is dawning awareness of Counterargument/Concession. One teacher asked if Paper 2.2 would not be strengthened with a counterargument; and in the next mention, another person stated that the author of Paper 2.4 had made only a weak attempt at counterargument. An interesting exchange occurred in the discussion of Paper 2.3 regarding rhetoric. One of the high school teachers posted her own score and comments on the paper, then downloaded the college commentary. After hearing some remarks about Aristotelian and Rogerian models of argument, she went back to discussion forum and entered this post:

> After the BWC commentary, I must ask (and in the process risk revealing my ignorance): What are Rogerian and Aristotelian methods?

In a show of support, another teacher responded immediately, "Right on." Later, a third teacher posted links to a couple Web pages where the terms were explained. It is likely that none of the high school teachers were familiar with these terms at the beginning of the study. In Round Three, there are six comments about counterargument, with 10 more in the last two Rounds. Two examples from the paper on teen parenting and one about the recycling paper are typical of the high school teachers' mention of Counterargument/Concession:

> This paper does not include any concession about the reality of teen parenting. For example, isn't it something like half of all teen mothers wind up on welfare? Also, I believe I read a statistic that indicated that children of teen parents are three times more likely to wind up in prison.

> The second paragraph on page two attempts to show the other side: that the movies really are heavily one-sided; again, the author uses her voice of experience—her best friend saga—to validate herself as someone who knows of what she speaks. However, this effort is feeble, since it ends with "happy grandparents."

> The author successfully addresses a complex topic with complex solutions and problems. He/she handles the opposition quite well towards the end of the paper.

Even in the second example above, in which the teacher does not use the terms, "concession," "counterargument," or "opposition," he is evaluating the essay in terms of its fairness in presenting both sides of the issue. In all three examples, the teachers not only show an awareness of the presence of Counterargument/Concession, but evince a positive valuing of the Dimension. By the end of the project, all six high school teachers had commented on Counterargument/Concession at least once, indicating that this feature of writing was now recognized and considered worthy of comment.

3) *Explicit Connection.* The college faculty dealt with the issues of clarity and coherence at length, although their comments did not use the terms "clear" or "clarity" often enough for it to be included as a Dimension, except in relation to the thesis. "Coherence," however, did appear quite often in relation to the argument, and it was identified as a separate Dimension. The college emphasis on coherence was frequently expressed in comments about how well writers made explicit connections between ideas and between an idea and the thesis, so much so that "Explicit Connection" emerged as a separate Dimension. It was defined as,

> *"The relationships between evidence and conclusions drawn from it; between various parts of the argument and each other; and between parts of the argument and the thesis must be clear. The writer does not leave readers to make inferences or connect relationships on their own."*

They often used the words "explicit," "relationship," "connection," and "link" in this context. The following examples are from the college commentaries on Paper 1.4 and Paper 2.4 [italics added]:

> To begin with...the introduction has a number of issues, and it's indicative of what we saw throughout this essay, which is the fact that *the writer rarely establishes links or connections* between things that he's saying, comments that he makes or statements that he makes.
>
> What's interesting to us, and becomes a concern as we read this paper, is that ... when you read the introductory paragraph, it's really a challenge to find a specific thesis. We find a sentence, "If we find a way to bring the price down it will help the economy because people will buy more items" Then we see the sentence, "If we use gas alternatives the price will go down and prevent global warning." So in the introduction we see the notion that if we can find a way to reduce the price of gasoline, it will have some economic impact. It may also have some environmental impact. Our problem is *that the relationship between the economy and the global warming and the gasoline is never explicitly stated in a comprehensive, complex thesis*. And because the general idea of the paper *is never specifically expressed in one clear statement,* then the direction of the paper suffers a bit.

The overall thrust of the college teachers' comments was that no particular sequencing of ideas was required as long as the writer's argument was both clear and coherent. Several of the papers they analyzed achieved clarity and coherence through articulation of relationships in spite of non-chronological or non-linear arrangement of ideas. A key element in producing clarity and coherence without relying on organization, according to their commentaries, was the explicit expression of the relationships, links, or connections among the various ideas in the paper.

The high school teachers were equally concerned about coherence, but they initially thought of it in relation to the 5-paragraph essay structure. In other words, they saw incoherence as a problem caused by failure to adhere to the five-paragraph form, and

they thought the solution was adherence to the form. Generally the form requires an introduction with a thesis statement in the last sentence, at least three supporting paragraphs, each of which begins with a topic sentence relating the paragraph to the thesis, and a conclusion that summarizes the argument and restates the thesis. These elements are mentioned frequently by the high school teachers in Round One:

> The writer seriously lacks topic sentences that would further his/her claim about robots serving as nurse assistants rather than replacements.
>
> The author should also be wary of over-citing and frequently beginning paragraphs with quotations and not topic sentences.
>
> A clear thesis is lacking in part 1. No topic sentences related to argument are present.
>
> The thesis is evident, but states the obvious in a very rudimentary way. This writer has clearly been trained in the 5-paragraph format.
>
> When I began reading, I thought, "Yes! Finally, a write[r] who uses a thesis statement in the first paragraph.

The first mention of the 5-paragraph form in the college commentaries came in Paper 1.3, when it was referred to it in a rather derogatory way:

> The thesis is directly stated in the in-your-face title and in the last sentence of the last paragraph, making this a delayed-thesis essay – and that's good. *But* – [sic]that scheme echoes the old five-paragraph formula essay plan which directs writers to state the thesis early and then repeat it again in different words in the concluding paragraph.

Although his tone of voice cannot be heard in the printed quotation, it is clearly disparaging in the recording. In the next paper, two of the six high school teachers bring up the 5-paragraph form. The second post is a response to the first:

> I really don't know why the 5 paragraph essay [sic] has become a bad outline to follow. It helps the writer to outline a course to follow with their

> information and, if followed, ensures that clear thesis statements are made and points are supported with evidence.
>
> Thank you! I also would like clarification on the "evil 5 paragraph essay." I teach this to my sophomores primarily to give them structure. I feel that even though this paper (1.4) was simple, it was easier to follow because of this structure. What is worse—a paper that is all over the place in regards to organization or one that follows a simple format?

The same two teachers comment again in the discussion on Paper 3.1:

> It was clear that this writer had a pretty good grasp on how to organize and how to document sources. The organization was very clear; however, I'm sure the BWC commentary will mention how it follows the dreaded five-paragraph essay format---- this concern is something that I'm trying to understand.
>
> I still need to be enlightened as to why the "5 paragraph essay" is dreaded!

At the time of this exchange, the BWC teacher who was following the discussion was writing several paragraphs to another teacher about the role of conventions in grading (another sensitive issue), and she did not respond to concerns about the Five-Paragraph Essay. The matter was not directly addressed by the BWC faculty in the rest of the project, and the issue of the Five-Paragraph Essay was never resolved to the satisfaction of the high school teachers. It was still bothersome to them in this Round Five exchange among three teachers:

> *Comment:*
>
> In light of the comments by BWC during the last couple of rounds, here's my million-dollar question:
>
> In high school, should a thesis statement be acceptable in the ninth or tenth paragraph of a writing? I most definitely expect to see it in the first paragraph with my sophomores (though I am more open about its placement with my juniors).

> Many of my tenth-graders have a terrible time organizing their written thoughts. My requiring their thesis statement to be in the first paragraph "forces" them to address their main point and THEN work at supporting it appropriately.
>
> This concern leads me to revisit someone's question (which I believe has remained unanswered) from earlier rounds: What IS wrong with the more formulaic five-paragraph format with the thesis in the introduction?
>
> *Response:*
>
> Thank you for asking this question again. I believe I was the one who posted the question about teaching our students the simple, formulaic essay. I too teach my sophomores to include the thesis in the introduction paragraph (typically the last sentence). Is this something that should be discouraged?
>
> *Response:*
>
> It seems to me that we HAVE to teach our students how to organize their thoughts, and the 5 paragraph formula helps to do this. I guess once students are in college they may be learning more sophisticated and creative writing styles, but a reader still wants to know fairly early in an essay what exactly is going to be addressed in that essay!

In spite of their defense of the 5-paragraph form, the high school teachers quickly began following the lead of the college faculty on the issue of Explicit Connections. The same teacher who was first to echo the college commentary on the Dimension of Counterargument and Concession was the first to express the need for ideas to link to the thesis, in Paper 1.4. She used the words "connection" and "tie," and never referred to 5-paragraph essay structure in this post [italics added]:

> He has tried to provide some historical context for this contemporary issue, but *there is little connection* to the cause and effect of its origins. The discussion of the barge, for example, is presented as an example, but does *not tie to the writer's thesis* nor does he ever *get to the point* about why the barge is even discussed. These "leaps" leave a writer disoriented.

Another teacher commented on the same problem in the next paper, 1.5. The problem she pointed out could as easily have been explained as a failure to follow the 5-paragraph formula with topic sentences required for each paragraph, but she eschewed that in favor of the "explicit connection" language of the college faculty [italics added]:

> Then a series of interesting questions and answers come along, but what have they to do with the first six paragraphs of the paper? After the end of the interview, no discussion *that would relate the interview back to the first half* of the paper occurs.

These two examples of the high school teachers forsaking reliance on the 5-paragraph essay structure to provide coherence are the only ones occurring in Round One, but the rate increased. In place of topic sentences, the teachers began using the words "connection, "tie," and "explicit" early in Round Two, as in the following examples. The first example wavered between the notion of expressing relationship explicitly and using topic sentences, but it ended with the more complex writing strategy of relating ideas to each other [italics added]:

> The discussion of urban gaming *needs (and can be) tied directly back to the thesis.* The subsequent paragraph also *needs a topic sentence.* Author might consider adding social bonding to the list of what makes a game successful. Or, *he needs to explicitly state* that the social bonding he exemplifies in the final paragraph is a product of the interaction component that the game companies intended.
>
> I like many of the *connections* made here and also like that the writer makes clear points on the complexities of the issue.
>
> The scenario at the beginning is an effective way to grab a reader's attention, but it ends abruptly *before it makes any connection* to the writer's point. There *is no transition* at all before the writer begins throwing in statistics and dates.
>
> For example: In paragraph 5, the last sentence, about pornography seems to come out of nowhere and *have no connection to the previous ideas.* In

> paragraph 6, the argument, I believe, is that students should be reading these controversial books with a teacher in order to understand the ideas better - but it is presented in a confusing way. In paragraph 7, the last sentence seems to be a repeat of a previous idea *but doesn't connect to the material* in that paragraph.

By the end of the project the teachers were regularly discussing explicit connections rather than the five-paragraph form. They also gradually decreased their comments about topic sentences, so that only three references to them occurred in Round Five, and one of those was a general comment not related to a particular paper.

Analysis of other Dimensions. Several of the other Dimensions showed increasing approximation of a single high school teacher's comments to the college commentaries, but appeared in the analysis of only one or two teachers rather than being representative of the entire group. Therefore, these Dimensions were not considered as indications of community formation. They do help explain *how* community formation occurred, however, so they are discussed in the individual narratives prepared for analysis of the second research question about how community formed.

During their commentaries about the three Dimensions discussed above, the college teachers never explained their views directly to the high school group or attempted to "correct" or "enlighten" their thinking. They simply engaged in the type of analysis of student writing that was familiar to them, and the high school teachers quickly picked up both the concepts and language.

Findings about the Perspectives on Writing

After the Dimensions of writing were coded, a second process of constant comparative coding was used to identify instances of dialog that indicate ways of talking or thinking *about* writing and the teaching of writing, as opposed to Features or

Dimensions of the writing itself. In this second coding process, eleven Perspectives on writing emerged, as noted in Table 5. Definitions of the eleven Perspectives are available in Appendix K, p. 187.

Table 5: Perspectives on Writing	
1	Awareness of Writer
2	Pedagogical Approach
3	Reader Response
4	Positive Focus
5	5-Paragraph Essay Challenges
6	Conventions Questioning
7	Stress of Teaching
8	Expressions of Disagreement
9	Introspective Comments
10	Expressions of Agreement
11	Questions to Group

In four of these eleven Perspectives a shift occurred by the end of the study that indicated the high school teachers thought and talked about writing in ways that were similar to the college teachers. I have labeled them 1) the Pedagogcial Approach; 2) Awareness of Writer; 3) Reader Response; and 4) Positive Focus. They are discussed in that order in the following sections.

1) Pedagogical Approach. The first of these four shifts was of the stance of the teacher in relation to the student writer. It was defined as, *"The teacher framed criticisms of the paper not in terms of good/bad or right/wrong, but in language that would be helpful to the writer in understanding his/her writing and in revising it."* Nearly all the comments by high school teachers in Round One projected the position of the teacher as an evaluator doing summary evaluation: the work of the student has been completed and the responsibility of the teacher is to pass judgment on a finished product. The college

faculty, by contrast, frequently took a more pedagogical approach to analysis of the student writing. Their comments consistently reflected a concern for how the paper could be improved and what revisions would strengthen it. In Paper 1.2 about televised sports, for example, the college commenter referred to the writer in the third person, but was thinking about what the student needs to do to improve the paper, almost as if speaking directly to the student:

> Continuing on with that argument, the author needs to establish why quality time can't be achieved by viewing televised sports, and then, tell us what makes watching televised sports any worse than any other program, like The Colbert Report, CNN (with "The Best Political Team on Television"), Bill Maher, American Idol, for example. What's the difference between watching television sports and watching those other programs? We have no idea.

Another example from the college commentary on Paper 1.4 illustrates that the faculty's concern is more about "how to fix it" than about "what grade it deserves."

> Again, if the writer had used his sources differently, he probably could have made a good point for his argument. If he had used some quotes from some of the sources and directly linked them to the argument, that would also have given some strength to his writing.

The first comment from a high school teacher that could be classified as Pedagogical Approach came in Paper 1.5, with this posting:

> This would be a time where I would like to dialog with the writer and ask them to expound on their purpose and how they tried to organize their ideas with support. It seems like a first draft.

This is the first indication of an attitude of *formative* assessment, in which the teacher was thinking not of what final judgment is appropriate, but how the assessment would serve the student's learning. It was the only such comment in Round One. In Round Two,

however, eight comments reflected a pedagogical approach. The number grew to 13 in Round Three, 15 in Round Four, and 23 in Round Five. Typical of the early Pedagogical comments were these examples:

> I would advise the writer to decide on a stance and avoid using the word "seems" so much!
>
> Suggestions for improvement: Explore the idea of "instant gratification" further – is it human nature? Work on improving the thesis. Research to strengthen the argument.

At one point in Round Three, several high school teachers expressed concern over the weight of the written conventions (spelling, grammar, punctuation, etc.) in grading. Their concern occurred against a backdrop (expressed later) that they felt a great deal of tension over the need to maintain high academic standards and to achieve high test scores on one hand, while meeting their principals' insistence that no student could be failed on the other. Much discussion ensued. In response, one of the college teachers described this model for grading:

> "No student in my class will receive a 4.0 on a paper if there are more than a few mechanical errors. Any paper riddle[d] with errors will have to be rewritten and the changes made before I will put a grade in the gradebook.

After this point in the discussion forum, revision became a frequent element in comments about papers. Here are some typical examples:

> With minimal clean up, I would then give it a 4.0.
>
> The paper had a few mechanical errors I would like to see corrected in a revision.
>
> The author does have a thesis, but needs to make it more complex. I would suggest modifying by saying how the companies have team[e]d up to keep people playing. The author keeps saying that interaction is the key

> component of a successful game. I would challenge the writer to show how a blend of interaction and the newest technologies keep people playing games. This combination of interaction and technology is latent in the paper, but the author needs to explicit[ly] say it and explain it when analyzing each game (video, DVD, etc.).

The teachers also began discussing pedagogy directly, as evidenced in these posts:

> I like the BWC commentator's advice to make copies of past student work and edit with current students.
>
> This might be another great paper to use as a teaching tool with our student to discuss how to come to a healthy balance of citations and author statements/analysis.

Finally, in Round Five, after reading a particularly deficient paper, one teacher asked the college faculty:

> Now, here is my question: How did you respond to this writer? Such glaring problems are far more serious than a misplaced comma. Where did you even begin?
>
> I wonder if it's appropriate in this project to ask each other about just HOW we go about reacting to our students' writings? For example, I do not correct mistakes (usually). I do make a check mark above any "problem" (misspelling, grammar, conventions) and expect a student to decide what the correct fix is. Peer editing has its place, but I have not ever really been able to employ it as successfully or as consistently as I would like.
>
> I always try to respond positively to a student's writing by including a "holistic" written comment, but then use shorter questions/comments in the margins to point out areas needing improvement. Unfortunately, high school students tend to simply search for a letter grade and gloss over any comments I make.

The college teacher monitoring the discussion responded to this post with several paragraphs, giving some practical suggestions about projecting past papers and editing as a class, and about how to make peer editing work. She stressed the importance of responding to the quality of coherence throughout the length of each paper. Perhaps more

importantly, she sympathized with the higher class loads of high school teachers that make personal conferencing almost impossible. Her suggestions were reflected in several posts in the remainder of Round Five, as the high school participants frequently discuss the pedagogical implications of their evaluative comments.

One of the most interesting – and unexpected – findings of this study is this change of attitude from summative assessment to formative assessment. Except for the single comment about conferring with a student, none of the comments in Round One reflected anything but the summative position. By the end of Round Five, a great number of comments indicated an awareness that assessment is meant to improve student writing, and people were thinking actively about what the authors of each piece must do in the revision, a step in the writing process that was not mentioned at all in Round One, but was taken for granted in Round Five.

2) Awareness of Writer. Along with a shift in attitude from summative assessor to pedagogical assessor came a shift in sensibility about student writers, from thinking about the paper as an isolated artifact to thinking about the paper as an extension of a real person. Awareness of Writer was defined as, *"The teacher thought actively about the writer behind the paper, not just noting what the person did wrong or does not know, but considering what the writer thought, wanted to achieve, was trying to accomplish, decided, and so on."* Throughout Rounds One and Two, the high school teachers treated each paper not only as a completed product to be graded, but as a product that had little relation to its author. They frequently referred to "the writer" or "the author" or "the student," but without much sensibility or feeling for an intellect struggling just behind the

paper. Nearly all of the comments in the early Rounds referred to the writer in cursory or even derogatory tones, as in these examples from Round One:

> The writer also did not know how to use a semi-colon, often using it where a colon should be used.
>
> He/she was trying to create an interesting topic, but fell short because of the aforementioned lack of thesis and disorganization.
>
> Otherwise, the writer seems to have a good grip on the subject and on documenting the research.
>
> The author should also be wary of over-citing and frequently beginning paragraphs with quotations and not topic sentences.
>
> The author needed to state clearly in the opening paragraph his/her argument that robots cannot replace human care but can certainly augment the patient's care, or something to that effect.

In such comments, the writer was referred to almost as a distant object whose faults needed correcting. Then in the beginning of Round Four, after the Pedagogical Approach had emerged, one teacher made this observation:

> While there was a thesis, I felt that the author wandered away from its point in a couple of places. Honestly, I feel that this could have been lessened had she written the paper in a more sequential way. This is a narrative . . . why did she not write about the tsunami earlier? That would have tied in with the "changing expectations" theme. Why did she NOT write about the trans-Atlantic tickets nearer the end? I understand that some points were to be interpreted as flashbacks, but this organization did not serve her well.

In this post, the teacher is no longer thinking about the text merely as an object, nor about the writer as a disembodied object, but is actively wondering about the student. There is a sensitivity to the presence of someone behind the text who made a choice and was trying to accomplish something, but fell short of the mark. This was followed by another post on the same paper:

> I thought that the writer's organizational choice was interesting. Although it seemed somewhat jumbled at first, I ended up liking the lack of chronology/ sequence. I think that it added to the overall voice of the paper, and through the details, the story kind of unfolded itself.

In the next paper, another teacher comments:

> Inexplicably, paragraph 11 launches into a pondering of Indian folklore. Why did the writer not begin with this point, if he felt inclined to include this? It certainly would have provided a much more interesting introduction to this piece that [sic] what exists now.

Next, about a paper entitled "What My Purse Means to Me," came this passage:

> It was evident that she wanted to discuss her purse, but this topic seemed to lead her into a new idea (her keys). Early on, it also seemed that she considered taking a feminist approach, but glossed over that as well. Again, I will point out that yet another writer is unsure of his/her purpose.

The same teacher showed awareness of the writer as a person in this Round Five comment on mechanics:

> When editing, the student needs to track the modifiers to the word they are associated with. I notice tracking has been scrupulously done in regard to subject/verb agreement, so I know the student is capable of this additional task.

Eleven comments of this type occurred in Round Five, ending with this one:

> From reading what the author had to say, we realize that he is a new college student with a lot to learn about mechanics, grammar, and proper essay writing (maybe he didn't get such an excellent education in the old building! just kidding). However, I did not grade him harshly on these flaws because I think he makes a lot of really good points in his discussion of a real issue that our state faces: outdated school buildings.

An interesting feature of these comments was that the teachers relied heavily on judgmental language early in the study, as in the examples cited from Round One quotations above:

> writer also did not know fell short

The author should also be wary The author needed to state clearly

By Round Five, the teachers use more verbs that signify some cognitive action on the part of the writer and nouns that name aspects of the writing process, as if they are describing how the writer faced choices and made decisions during the act of writing, as typified in these phrases from the quotations above:

the author wandered away

why did she not write

the writer's organizational choice

paragraph 11 launches into

she wanted to discuss

she considered taking

glossed over that

tracking has been scrupulously done

The shift from thinking of the writer as a faulty, isolated entity to thinking of the writer as a processer of ideas who is making choices was accompanied by a parallel shift from thinking about writing as an artifact to thinking about writing as a living result of cognitive processes that involve choices, decisions, and consequences.

3) Reader Response. In addition to the emergence of the Pedagogical Approach and Awareness of Writer, another Perspective arose late in the project. Reader Response was defined as,

> *"Consistent with the work of Louise Phelps (1989), the teacher approached student writing from the perspective of a reader, noting what makes sense, is easy to follow, is clear, is convincing, and so on, rather than as an evaluator judging how right or wrong the writing is."*

It is interesting to note that none of the college commentaries explained Reader Response theory or even mentioned it directly. Rather, the high school teachers seemed to have "picked it up" intuitively by adopting the model offered frequently by the college faculty. Developed by Louise Phelps, reader response theory contends that teachers are most helpful giving feedback to students when they read and respond as readers instead of as evaluators (Huot, 2002). The most memorable example of Reader Response in the college commentaries (and the one that provided the deepest chuckle of the project) occurred in Paper 3.4 on the subject of government-provided health care. The college teacher was explaining the dangers of using the "impersonal you" in writing:

> When you use sarcasm and when you use "you," it's so easy to turn a person off. Even if I agree that we need universal health care, I'm reading along and all of a sudden I come to this sentence that says, "We have been taught from a very young age that it is a dog eat dog world and the only way to survive is to look out for numero uno, you." First of all, I'm offended by the cliché. And then, I'm like, no I wasn't taught that. And all of a sudden I'm arguing about his characterization of *me* rather than the topic. And you don't want your reader to go into that zone, you want to keep the topic on focus. And I'm less persuaded because he's just accused my parents of raising me up wrong. And that's not true. Don't insult my mother!

It is doubtful that any students in her class would forget such a lesson on the "impersonal you," because the instructor explained so vividly how she, as a reader, reacted negatively to the language and so found the entire argument less persuasive. If she had pointed to a rule in a stylebook, the lesson might have been forgettable. It was her response as one human being reading the words of another that was convincing. Reader Response was not

universal among the college faculty who participated in this project, but because the commentaries were always developed in pairs, there was always at least one person who modeled Reader Response in each of the recorded commentaries.

After the teachers began discussing assessment of papers in a pedagogical mode and became more aware of the student behind the paper, it was as if they also became aware of themselves. Three times in Round Four and 11 times in Round Five, comments were framed in a way that acknowledged the presence of a reader: in other words, the teachers shifted their role from that of judge to that of reader, a person on the receiving end of a communication who brings experiences, values, and feelings to the act of reading, and whose need for clarity and direction is met more or less well by a writer leading the way through the text. The first instance of a Reader Response comment occurred in Paper 4.3 (most italics are added throughout these quotations; the only original italics are noted):

> In the fourth line, "Being a mother you..." is not punctuated correctly and is too wordy: the plural *Mothers* [italics original] simplifies without loss of meaning. Using "It" without a clear antecedent is too conversational for formal writing. *I am still onboard*, though, and *think I understand the intent* of the paper. *I am looking forward* to experiencing the purse's role as an extension of a woman's femininity. Unhappily, *I am not able to find* any trace of this controlling idea throughout the rest of the paper.

After three sentences in the evaluative mode, the commenter of this paragraph indicated his own reactions to what he read, not just as a teacher, but as a person trying to follow the argument. He was energetically trying to follow the writer through the thicket of prose, and he expressed the positive emotion of "looking forward," followed by an emotion of disappointment as the writer fails to deliver on his expectations. This comment showed awareness of self as a reader as well as a teacher. Another participant

made the following comments about a reader's need for clarity generally, immediately linking the thought to her own confusion about the writer's direction with the topic. The paper is about teen pregnancy.

> *I found it confusing* and contradictory, as you did, that paragraph 6 says "there will be a decrease", and then paragraph 7 says "So maybe there will still be a rise . . . " and again in paragraph 10 "The [n]umber of teens having babies is going to rise." ... The writer needs to first figure out his/her opinion on all of this and then communicate that clearly to the reader...When I began the paper, *I thought it was going* in a whole different direction from where it ended!

Another participant expressed the same confusion about the writer's stance in the same paper. At various points he was aware of the fact that he was guessing about which side the writer would come down on:

> That the taboo against having babies has been lifted is clear enough; nearly all the accompanying fears, shames, and excommunication from polite society have been deflated. *I am still not sure* whether the author approves of this or not. . . The fact that she overstates her argument on page two *continues my belief she is* supportive...The paper ends with a hearty endorsement that "pregnancy is cool and exiting at any age." At least *I finally get* my long-awaited judgment.

In these responses, the teacher had taken on the role of a reader being led by a writer. Another example occurred in Paper 5.3, about the presence of discarded medicines in waterways. It was one of the more successful papers in the project. In one paragraph, however, the writer "lost" the reader. Again, the problem was expressed not so much as a fault of poor writing as awareness of the teacher's own need for the writer to provide clear direction and signals:

> At this point, *I realized that I was confused* about the writer's intended point (and audience?) Was he/she trying to point out danger to fish or humans?

The same participant brought her own prior experience into the reading of Paper 5.4, acknowledging that it influenced her to react positively to one idea and to expect it would lead to a particular kind of scenario, which was different from the writer's scenario. She also acknowledged a negative emotional reaction to the tone of one sentence:

> Para. 2--Idea about safety net *coincides with my own thinking*, so *I found* it particularly gratifying. However, the jump to rape, drug abuse, and wild partying *was too big for me. When I think* cell phones and safety nets, I think "flat tire" or "out of gas." *I thought it valid* that the author questions the strength of that safety net, but *was put off* by the almost-condescending tone in this sentence: "Let us be practical and assume the reality of teen behavior."

These few examples do not indicate that participating teachers had developed fully into practitioners of the Reader Response theory, of course, but since no comments from the first three Rounds could be coded in this Perspective and fourteen appear in the last two Rounds, it is reasonable to conclude that at least the seed had been planted, and that the impetus for the emergence of Reader Response comments among the high school teachers was the model presented by the college faculty.

4) Positive Focus. When I began reading the posts of the high school teachers in Round One, I was taken aback by their overall negative quality, which mirrors the "fault finding approach" described by Virginia Chappell (1991) and others. Many teachers read student papers by mentally comparing it with an ideal paper, marking and commenting at each point where the student work deviates from the ideal. The result is a "fault-finding summative evaluation of student writing that makes grades, their bestowal and receipt, so distasteful" (p. 55). Of course, sometimes student writing does seem to call for that approach. One teacher began her comment on Paper 1.4 by saying, "I must really work hard at finding much positive in this paper, but here goes."

By contrast, the college teachers nearly always began their commentaries with positive points and often cited other strengths of the paper in the ensuing analysis. Even the worst papers received some acknowledgement of accomplishment from the college faculty, including the one they excoriated most heavily and scored with a 0.0. Here is a transcription of the audio portion of the first two slides of that paper. The first speaker even manages to inject a positive tone in her voice, and the second speaker also makes a stab at an even tone.

> We believe the topic has a lot of potential. It's very timely and important. There was an article in the news just recently about some teenage girls who supposedly had some kind of pregnancy pact. They all wanted to be just like Jamie Lynn Spears, so...important, important topic!
> Again, from the social perspective, the writer does make some good points, although they're hidden among her technical problems, and her argument *could* have been somewhat well constructed. However, her thesis has issues, and it just doesn't all fall together in a logically formatted argument, but she does start out with a few redeeming qualities.

The following slide went on to list the main points of the writer's argument in a neutral and uncritical manner.

Two of the high school teachers were inclined to begin their posts in Round One with a general positive statement, but the others usually started abruptly by listing what they found wrong in the paper. In the middle of the project, on Paper 3.4, one teacher acknowledged the negative tone of her own comments, and identified how important the lack of context was when thinking about student writing from a pedagogical point of view.

> The above paragraph brings up another idea I've been struggling with throughout the process of scoring and commenting on these papers. When I usually am grading papers, it is not done in the vacuum that this project presents. I know my students and how they think and write. I have a sense of what is an improvement in their writing and when a student is just making the same old mistakes again. With these papers, there is no context

> for us; we don't even know what the assignment is, and I feel like I'm being really harsh on these papers, when in fact this may be huge strides in writing and thinking for some of these writers!

At the beginning of Round Five, the same person described her approach to paper grading this way:

> I always try to respond positively to a student's writing by including a "holistic" written comment, but then use shorter questions/comments in the margins to point out areas needing improvement. Unfortunately, high school students tend to simply search for a letter grade and gloss over any comments I make.

One of the college faculty added this comment:

> Responding holistically in a non-judgmental manner goes without saying, and I am fairly sure we all do that, anyway.

For the remainder of Round Five, five of the teachers seemed to make a more concerted effort to comment on the positives of the paper, although one teacher approached positive statements only by agreeing with earlier comments. The numerical counts of positive statements do not appreciably differ from Round One (21) to Round Five (29), partly because teachers began agreeing with other each so frequently, and partly because the number of positive comments varied, quite naturally, by the quality of the papers. A qualitative difference can be discerned in the placement and sense of comments between Rounds One and Five, however. By the end of the study, they are more frequently embedded in the detailed analysis rather than appearing only at the beginning of the post, and they more frequently point to specific features or traits rather than expressing a general "good job!" sentiment. Here are some examples of Positive Focus comments from Round Five posts:

> Paragraph 2: Great quote to start...
>
> Paragraph 3: Discussion of media is a very valid direction to take here...

The writer seems thoroughly interested and personally affected by this topic.

The paper possesses a reasonable degree of integrity, which is often lacking when students want to prove a point.

Para. 7--Writer continues on a good path by discussing the second half of his/her thesis.

By the end of the project, all but one of the high school teachers had begun analyzing the papers in a more balanced manner, noting the strengths of a paper as well as its faults, and commenting more often in neutral, descriptive, nonjudgmental language rather than the fault-finding mode of the first Round.

Summary. There is considerable evidence in both the quantitative and qualitative analyses that high school teachers began commenting on the papers in the manner of the college teachers in several respects. First, the correlations of interrater reliability improved rapidly at the beginning of the study for most participants, and generally continued showing improvement through Round Four. The correlations for the mean of the high school teachers' scores compared with the college teachers' scores were 0.94 for Round Four and 0.88 for Round Five, well above the standard of 0.80 for the testing industry (Elliot, 2003; Huot, 2002). Compared with a correlation of 0.19 for mean scores in Round One, the high interrater reliability is a strong indicator of the growth of interpretative community.

The evidence is most apparent, however, when comments posted at the beginning of the project are compared with comments posted in the final Round. One major outcome is that high school teachers began commenting extensively on the particulars of the students' arguments, pointing out faults and strengths of logic and evidence in great detail, as if they had entered the world of the argument themselves. They also came to

recognize the presence of Rogerian Counterargument and Concession and adopted it as a positive value in composition, and they began thinking about coherence as an issue of connections and relationships rather than as a matter of following the five-paragraph form. These three developments could be helpful for the high school teachers in becoming more effective in terms of preparing their students for college composition at Blue Water College and possibly other institutions as well, since it is likely that these Dimensions of writing are taught and valued in many first-year composition programs. Together, these three developments provide solid evidence of the formation of interpretative community among the teachers.

Perhaps even more important than becoming increasingly adept at handling issues of argument in student writing are the shifts that occurred in the teachers' manner of response to student writing. They began taking an approach to assessment that was infused with concern for how they would utilize the assessments in the practice of teaching, with a focus on student learning, and their comments increasingly pointed out strengths of the writing rather than taking the fault-finding approach. In the process, the high school teachers became more aware of the student as a writer and of themselves as readers.

By themselves, these developments, however valuable, do not indicate any formation of interpretive community. It is only when initial and final responses of high school teachers are compared with the college commentaries that formation of interpretative community is seen. The consistent presence in the college commentaries of these three Dimensions and four Perspectives, and the increasing reference and use of them among the high school teachers does indicate that the teachers began "seeing"

things, in Fish's sense, in similar ways. Furthermore, the changes in their comments are not limited to surface judgments such as, "This paper has too many errors to get a 3.5; the right score is 3.0," but point to deeper shifts in consciousness, attitude, and approach. The finding from qualitative analysis that interpretative community has formed is bolstered by rates of interrater reliability increasing from low to high over the course of the study (see pp. 52-53). Therefore, it is reasonable to conclude in answer to the first research question that an interpretive community did indeed form among the high school and college teachers through their participation in the online assessment community.

Question 2: How Did Interpretative Community Develop?

Analysis of the Group as a Whole

Evidence about how interpretative community developed was sought from the analysis of the discussion forums in the second process of coding that identified eleven Perspectives, or ways of talking and thinking about student writing. In addition to the four Perspectives described immediately above (Pedagogical Approach, Awareness of Writer, Reader Response, and Positive Focus) and Expressions of Agreement mentioned earlier, six other Perspectives were identified by the constant comparison method. They are listed in Table 6, along with the number of comments for each one in each Round. (Definitions are available in Appendix K, p.187.)

Table 6: Perspectives Mentioned by Round

Perspective	Round 1	Round 2	Round 3	Round 4	Round 5
Pedagogical Approach	1	8	13	15	23
Awareness of Writer	0	0	1	6	11
Reader Response	0	0	0	3	11
Positive Focus	21	27	13	13	29
Role of 5-Paragraph Essay	8	1	2	1	5

Perspective	Round 1	Round 2	Round 3	Round 4	Round 5
Conventions of Written English	0	5	6	9	1
Stress of Teaching	3	2	1	2	2
Expressions of Agreement	5	17	11	18	21
Expressions of Disagreement	1	4	1	6	5
Introspective Comments	2	2	3	2	5
Questions to Group	1	5	2	3	6

It seemed reasonable to assume that items such as Introspective Comments, Expressions of Agreement, and Expressions of Disagreement would be relevant to the formation of interpretative community. I read all of these comments in their context for the purpose of identifying specific words, phrases, or sentences that would indicate reasons for change in teachers' thinking about the Dimensions or scoring. I especially looked at the discussion just prior to the coded comment for material that would explain the cause of a Perspective comment, and I read the discussion immediately after each one to locate any specific turning points in the teacher's thoughts. This procedure was not fruitful. Although there was a considerable amount of potential evidence, it was weak and required too much interpretation to be trustworthy.

For example, a great deal of dialog occurred among participants in Round One over the role of the five-paragraph essay, and in the middle rounds over the role of the Conventions of Written English. These exchanges of ideas were often accompanied by Expressions of Agreement and other supportive comments, and I was tempted to conclude that Expressions of Agreement were a key element in the formation of community. It may be that they are, but the evidence was "slippery:" it could not be connected directly at a moment in time with specific turning points in teacher thinking. Even with the Introspective Comments, nothing emerged from data in the immediate

context that provided a supportable explanation or theory about how interpretive community formed. The eleven Perspectives listed above were present intermittently during the course of the project, but the attempt to connect them with other comments immediately preceding or following them provided no direct, clear evidence of why or how interpretive community developed.

A similar effort to explain how community formed by rereading the interviews was similarly unsuccessful. The interviews contained many instances of introspection, thoughts about interactions with other participants, and reflections about the experience that could possibly be interpreted as causes for the formation of community. The links were weak, however, and required too much interpretation to be trustworthy. Solid theoretical explanations for formation of interpretative community observed in this study come from Wenger's (1998) work on communities of practice and Bruffee's (1999) work on collaborative learning, and I will discuss this finding in the light of some of the literature in the Conclusions chapter. No explanation of how community developed emerged from this analysis, however.

Individual Narratives

The best evidence for explaining how interpretive community formed came from examining each participant's comments separately from the entire discussion. After considering general patterns in the record of all participants, it made sense to look at each participant's record individually, since each one started in a different place and moved in a different direction to arrive at the central place of interpretative community. Those individual records coalesced as narratives, which are presented below.

Viewed individually, patterns in comments indicate that change was more pronounced in some participants than in others. Interestingly, the degree of change corresponds very closely to the amount of time participants spent on the project and to the length and detail of their comments. The two participants who spent the most time on the project and commented in the most detail are those whose records indicate the greatest amount of change: FR, whose total of 252 comments on the Dimensions far outnumbered the second participant's total of 146, and RR, whose total word count of 5,560 almost doubled that of the next participant, are the first two participants listed in several of the analyses described earlier in this report. Conversely, LL, whose comments (77) and word count (1042) were the lowest in both categories, showed the least change in several of these analyses. Whether this correspondence indicates a cause-and-effect relationship, a simple association, or merely that a greater amount of material was available for analysis, is unclear. It seems reasonable, however, that greater amounts of cognitive effort, practice analyzing student work, dialog with others, and reflection would result in greater changes in a teacher's analysis, a relationship that would make an interesting study for the future.

As with the group analysis, the comments of each participant were studied for evidence of what precipitated change. Four of the Perspectives that developed from the second coding process were especially explored for this purpose: Introspection, Expressions of Agreement, Expressions of Disagreement, and Questions to Group. In a few cases, Introspective Comments seemed to accompany an increase in attention to one of the Dimensions or one of the other Perspectives, but the connection was not strong, and none of the other Perspectives stood out in relation to changes in the participants'

comments. In short, I was unable to find any specific evidence in the form of Perspective comments that linked directly and temporally to changes in teacher comments that would explain *how* change occurred, even within these individual narratives.

Nevertheless, the individual narratives illuminate a fundamental part of the process of learning and growth: reacculturating to a new way of thinking without completely rejecting or abandoning previously-held values, but rather nesting them within the values of the new, larger community. Each of the following stories is different, because each participant started with a mindset about student writing that was substantially different from the others, and different to an even greater degree from that of the college teachers. The narratives provide a window into the thinking of the teachers, first by describing how distinctive they were from each other at the beginning, and second by explaining how their initial values and perspectives were reconciled with new ways of thinking presented by others in the group and by the college faculty.

The following sections each relate a narrative for one of the six participants considered separately. In all cases, material from the initial interview and Round One discussion forums helped explain the teacher's concerns and perceptions about writing at the beginning of the study. Although the middle three weeks of the project and the midproject interviews were studied for anything that would indicate how change occurred, changes were gradual and incremental, so that only contrasts between beginning and ending comments were salient. Material from the Round Five discussion forums and end-of-project interviews, compared with initial material, shows the degree to which the participant had changed in perception and valuing of student writing and how they resolved old and new ways of thinking.

FR's story. FR was noticeably the most engaged and dedicated of the six teachers. She was frequently the first to score new papers and post on the discussion forum. The number of dimensions upon which she commented over the course of the project was 252, compared with 146 for the next highest person. Although her online time logged only slightly more than the 20 hours required, it was evident from her posts that she had spent far more time than that offline preparing her analyses of papers and comments. Her interviews ran well over an hour, and there was often an earnest, almost striving, quality in her voice. She expressed apprehension for her students, their education, and the whole educational system frequently. Throughout the project, FR's analysis of papers was thoughtful, thorough, detailed, and amply illustrated with quoted material. She demonstrated a high level of knowledge about the writing process initially, and she was often the first to pick up on ideas introduced by the college instructors.

FR had been trained in the Six Traits system of teaching writing, which had been formally adopted by her school district, and she used it as a guide for both teaching and grading. Six Traits is widely accepted in U.S. K-12 English/Language Arts. The six traits are Ideas, Organization, Voice, Sentence Fluency, Word Choice, and Conventions. It is occasionally used at the high school level, but was originally developed for elementary schools and is used primarily at that level (Culham, 2004). In her first interview, she said of the system, "I love it." She expressed dismay that teachers of other subjects in her school do not support writing instruction "across the curriculum" by assigning writing and using the Six Traits to grade it, as advocated by many authorities. She said the Six Traits are very consistent with the ACT Writing Test (required by the State), yet upon further reflection noted that the ACT rubric does not contain Voice and never uses the

word "thesis," but instead tells students to "take a stand." She feels pressure to prepare her students to do well on both the Six Traits and the ACT. Early in the first interview, she asked what rubric we would be using, and she was speechless for a moment when she heard that the project would not require one, but that she could use a rubric if she liked.

Her first post on the discussion forum begins, "I am using the 6 traits holistic rubric, as that is what I have used for years." On the second paper her commentary began with, "[T]here is a thesis in there somewhere, but I had to read and then re-read to realize what it was." This comment reflects the Six Traits emphasis on clear, focused ideas and a narrow, manageable topic. The second part of the post referred to the paper's Organization, and then she commented on Word Choice, Voice, and Conventions, hitting all the Six Traits in order except for Sentence Fluency. Her responses to papers follow the Six Traits pattern generally throughout Rounds One and Two.

In Round Three, FR began commenting on papers paragraph by paragraph, usually numbering them, and picking out whatever features seemed salient in that paragraph, and largely ignoring the outline of the Six Traits. Her comments on Paper 3.2 are typical of this Round:

> Para. 1--Two excellent opportunities for thesis statements, but it was unclear to me which one he intended. Parenthetical formatting is, of course, incorrect. He does offer some concession in this paragraph. Why does he say that censorship "is indeed very widespread" and then not support that claim with specifics?
>
> Para. 2--Isn't Salinger spelled with just one L? I'd have to look it up to remember.
>
> Para. 3--This paragraph amounts to a list. No substance, just a recitation.
>
> Para. 4--The writer is trying to implement transition words like "however," which is a positive. HOWEVER, he inexplicably mentions

> then that "pornography should be kept out of schools." Why attempt to weave together ideas and then tack on something that is off-topic? (By the way, what did he mean "and they are just fine"?--another off-topic observation with no support).
>
> Para. 5--Again, I interpret this as a list. Too many ideas here to provide any cohesiveness (misinterpretation, kids choosing to NOT read, require students to read, better to read in a classroom setting).
>
> Para. 6--This paragraph could have been focused on the dangers of censorship, but it was unclear to me.
>
> Para. 8--I like his observation that "to be against something," . . . one must "have knowledge of it."
>
> I'd say this is a good first draft about a very interesting topic.

At the beginning of Round Four, there was also a change in the tone of her comments. She seemed less concerned with evaluating specific features in a particular order and began thinking about the process of writing from the writer's point of view, as in this comment on Paper 4.1:

> I paid little attention to the punctuation problems, as I felt there were more serious concerns. While there was a thesis, I felt that the author wandered away from its point in a couple of places. Honestly, I feel that this could have been lessened had she written the paper in a more sequential way. This is a narrative . . . why did she not write about the tsunami earlier? That would have tied in with the "changing expectations" theme. Why did she NOT write about the trans-Atlantic tickets nearer the end? I understand that some points were to be interpreted as flashbacks, but this organization did not serve her well.

Another new line of thought appeared in FR's Round Four posts: in addition to evaluating the paper, she began thinking about how she would respond to the student pedagogically. In other words, instead of limiting herself to judgments for the purpose of

summative assessment, she began articulating thoughts of more formative assessment, as in this comment on Paper 4.2:

> I would probably offer this paper back without a grade and ask for a revision before I evaluated it.

She directly voiced a concern for the pedagogical aspects of assessment in her post on Paper 5.1:

> I wonder if it's appropriate in this project to ask each other about just HOW we go about reacting to our students' writings? For example, I do not correct mistakes (usually). I do make a check mark above any "problem" (misspelling, grammar, conventions) and expect a student to decide what the correct fix is. Peer editing has its place, but I have not ever really been able to employ it as successfully or as consistently as I would like.

By the end of the project, FR had not abandoned the Six Traits Method: she still commented frequently on all those qualities of writing. She had embedded what she learned from that system, however, into a larger, more complex perspective about how to look at student writing. In this commentary on Paper 5.3, elements of the Six Traits are italicized to distinguish them from the broader analysis into which they had been incorporated:

> *I felt that the transitions and supporting examples were good. There was most definitely a thesis, to which the author returned at the conclusion.* I, too, enjoyed the scenario focused on "Century Bob." *Conventions were good, word choice was fine. Sentence structure good* (though the bulk of this paper was someone else's writing).
>
> In the introduction, the writer indicates that people in nursing homes would be in excruciating pain were it not for their meds. That's not necessarily true for many and so I found that an overgeneralization. I felt that the USGS information probably should have been cited.

Para. 2--Though the information is well handled, too much is cited here (in my estimation). Where is the writer's input on this issue? Or is it acceptable that he/she includes only outside research?

Para. 3--Author switches to fish, but makes no reference back to people. At this point, I realized that I was confused about the writer's intended point (and audience?). Was he/she trying to point out danger to fish or humans? (After all, he/she began by mentioning Century Bob at the nursing home. See my note re: paragraph 9--E. coli vs. Rx drugs) The sentence about feminized fish needs a cite. His/her last sentence in this paragraph is unnecessary or vague or both: "[I]t is not just an issue of accumulation of drug concentration harming wildlife, but the direct consequences of poisoning."

Para. 4--*Sentence one is a run-on. There is a comma missing from* "[A]s Andrew Johnson . . . says "We don't have enough evidence . . . ". Finally, I feel that too much is cited (though it flows well) and the author is simply compiling outside research without contributing any fresh thought.

Para. 5--I would have checked on this if it were my own student's paper: he/she mentions drug-resistant viruses, but I thought that E. coli was bacterial. I was baffled and will check it after I post this.

Para. 6--*Writer begins to come full circle by returning to his/her thesis (good!).* He/she probably should have explained what it means to "perform Extended Producer Responsibility." *(Why the random capitalization?)* I also found it a leap in logic to assume that returning unused meds to a licensed individual would prevent misconduct or abuse.

Para. 7--*Writer continues on a good path by discussing the second half of his/her thesis (education).* Probably should have explained the SMARxT acronym. The parenthetical was oddly formatted: ("Drugging Our").

Para. 8--The concession is present, but oddly located. *Writer undermines thesis when he/she states there is "only so much that can be done" . . . NOT the most persuasive observation one could make. The last sentence of this paragraph is very awkward: "stopping the flushing . . . and instead disposing of them."*

Para. 9--*Writer's point about defects in fish foreshadowing human defects should have been made in his/her introductory paragraph. That small reorganization would have definitely helped in the writing/logic department.* It is in this paragraph that he/she seems VERY unclear: is the topic E. coli in water or Rx drugs in water? For this reason ONLY, I found the reference to Century Bob a little awkward.

> Works cited very thorough, well formatted, alphabetized. However, I felt that an 8-page paper should not have as many as the 13 sources listed here. What does everyone else think?

Six Traits is a useful system for helping teachers deal with writing, but it is limited for preparing students for college-level writing. FR was able to integrate the familiar traits into a more complex scheme of analyzing student writing by the end of the five weeks.

RR's story. RR logged far more time online (35 hours) than the next participant (28.5 hours), but he said that part of the reason for that is that he types slowly. Part of it certainly was also that he provided detailed information, not only about errors of grammar and style, but about the source of the errors. RR had an appreciably more sophisticated understanding of grammar and style than the other five teachers, and he used this knowledge to analyze student writing extensively. This post from Paper 2.3 is typical of his work throughout the study:

> A superfluous comma exists in the first sentence, which tells me the writer doesn't know what a compound sentence is. In the purpose section, the nonrestrictive *which* is used, without a pair of commas, when the restrictive *that* is really needed. The use of a single comma just adds to the confusion and suggests the student does not know when to use a pair of commas to set off parenthetical interruptions or the difference between restrictive and nonrestrictive adjective clauses. In the audience portion, I question the word choice "holds" as well as the logic behind the sentence "Many people feel...", which sounds like a non sequitur to me. "Enthused" is used instead of enthusiastic.

In the initial interview, RR provided a detailed explanation of the influences that had shaped his interest in language and development as a teacher of English, from his mother to several teachers from junior high school through college. Several of these teachers had inculcated deeply-held values about writing, and at many points in the study

he seemed caught between the revered values of correctness and control of language he had learned as a student and the comparatively casual attitude he sees among his own students. He continues to read about language and teaching: he had just reread *The Practical Stylist* by Sheridan Baker earlier in the summer. He had also been strongly influenced by a workshop that related knowledge of grammar to reading comprehension. "It's all right there," he said in the interview. "If you just know the grammar, it's the key to everything."

RR's method of assessing student writing became clear early in Round One. His first comments are always about the topic and the thesis, and then he checks the conclusion to make sure it reflects the writer's original intent, as in his posts on the first three papers:

> 1.1
>
> The title serves no purpose and therefore is misleading; no thesis is stated in the intro paragraph... A weak opening paragraph. By skipping to the conclusion, I can see the thesis is finally stated.
>
> 1.2
>
> I liked the attention-getting title and opening; I would like to see the thesis appear in the opening paragraph; a thesis does emerge and the conclusion works for me.
>
> 1.3
>
> I feel the paper relies on the title for its direction; I really like to see a thesis statement in student writing; the thesis is implied, but is never completely stated.

Next he checks the supporting paragraphs for both support of argument and mechanics. RR justified his formulaic approach to grading papers even before he finished posting

comments on the first paper, and he made clear that his first priority was preparing students to score well in the state-mandated tests:

> I have too many papers to grade: I cannot be searching for key elements, any more than the persons who grade ACT or state essays—each paper gets a few minutes; that's it. (Spreading the thesis throughout the second paragraph is a good practice in the hands of a skillful writer, I don't question that.) My reading audience is the ACT and state, not college teachers. I know my readers will first look for a thesis; next, they will look at the conclusion to see if the paper even finishes the way it started. After that, they will look for support in the body paragraphs, scan mechanics in a broad way (intro material—comma; interrupters—pair of commas; compound sentences—comma or semicolon; and material tacked on the end—comma or not), and notice obvious shortcomings, such as *there is, you,* agreement, and pronouns without clear antecedent references. Whatever else they notice will be included in their evaluation: voice and style can really score heavy points when they stand out, for instance, but they must stand out.

RR's use of the state-test formula remains in use throughout the project: he usually comments on the topic, thesis, conclusion, and then, at length, on punctuation. By Round Five, however, RR begins his posts with a different approach, and the elements of the grading formula have been embedded in a larger context. Here are the first paragraphs of his last three posts:

> 5.3
>
> I was impressed by the overall maturity of the argument. I spent a great deal of time commenting on this paper already, only to have my comments not post. My remarks were regarding style, so nothing of any real consequence was lost. The student does very well already without any further comment from me.
>
> 5.4
>
> I will temper my enthusiasm for this paper simply because I like the idea and believe a good paper will emerge once editing occurs—I made this error earlier with the teen pregnancy paper. The entire business of self gratification needs [to be] dropped; immediate reinforcement is a better

> substitute since it is more defensible. Even so, the thesis needs [to be] broadened to include the student's scope of discussion . . . I like the energy of the paper; certainly the author bites off more than what is chewed, but this writer has something to say versus merely completing an assignment.
>
> 5.5
>
> This paper is not good, but the person who wrote it probably values it, and will be disappointed with a low C; I am speaking of the regular classroom student... This paper is very much a rough draft that would fit the description of using the overwriting technique. Still, the student feels he/she is done, when the editing process has not begun. At this point, I am glad I at least have a paper to work with, since several in the class will not even make this much effort.

As with FR, RR has not abandoned the scoring scheme he started with, but it is now one element of a larger scheme that considers the growth of the writer and the complexity of argument, and he does not rely exclusively on a formulaic response to student writing.

DQ's Story. DQ had taught in many different types of schools and named a variety of influences in her teaching. The Catholic school in which she started her career took a particular approach to writing that she still uses. She was quite concerned with students' ability to cite sources using MLA style in both in-text citations and the Works Cited page, and since the Works Cited pages were unavailable in the first few papers, she expressed some frustration and said she felt she was not able to assess the papers well. Other teachers, including the college faculty, also felt the Works Cited page was an important element in the consideration of a grade, but its absence appeared to destabilize DQ's whole approach to student papers.

1.1

> I teach my 10th, 11th and 12th graders MLA style, so they certainly should know how to cite sources in a Works Cited page and use parenthetical citation correctly. The lack of a Works Cited page is a real problem, because I had nothing to verify the various sources referred to in the paper.

1.3

> [I] agree with much of what Fluffy Ruffles says, though I believe paragraph 5 has only the one parenthetical citation at the end of the paragraph because all of the facts from that paragraph are from that same source. However, the period should go after the citation, not before.
>
> The lack of works cited pages is a big omision [sic]. Otherwise, the writer seems to have a good grip on the subject and on documenting the research.

The Feature of Citations included two Dimensions, 1) correctness of formatting in MLA style, and 2) whether a one-to-one correspondence existed between sources cited in-text and sources listed in the Works Cited page. DQ often commented on these dimensions either first or second in her posts, and they seemed to influence her overall judgment of the paper more than they did other teachers. She expressed appreciation for the inclusion of a Works Cited pages with the later papers, but her attention to these Dimensions tapered off, with only two comments in Round Three and one each in Round Four and Round Five. Her posts on Papers 5.4 and 5.5 focus far more on specific points of the argument. In addition, they show elements of the Pedagogical Approach and Reader Response, which are italicized in the following examples. (DQ's post on Paper 5.3 is omitted because, although it was about citations, it had been written in response to another teacher's question.)

5.4

I think I am in the minority here about the actual subject. *I found* the writer to be offensively simplistic and to overgeneralize a great deal. In paragraph 4 (page 3), the writer uses "(i)t is a well known fact" in two sentences. In both of these cases, the writer then connects the "well known fact" to something that does NOT logically follow.

Many of you noted that the "instant gratification" idea wasn't very closely followed throughout the paper. I found particularly offensive the idea that cell phones will limit today's teenagers' problem-solving abilities. *I have to admit that when I was 16 (eons ago) that I solved my car problem* the same way today's "instant gratification" teenagers do - except I had to find a pay phone to call my parents for help. *I don't think that* the lack of a cell phone gave me better problem-solving abilities. I have managed to somehow grow up and become a competent problem solver, and I'm fairly certain that my sons will someday be competent problem solvers as well. Actually, they are sometimes more creative and able than I was at their ages. I am not saying that I like cell phones, and I see many problems generated by them, but I don't think that there is a valid connection between having a cell phone and being a poor problem-solver. (I think that teens in general, then and now, are just poor problem-solvers - probably due to lack of life experiences!)

5.5

I did not feel that the author actually directly addressed the issue of how facilities can positively or adversely affect a student's education. I felt that some good ideas were raised, but nothing concrete was actually written *to convince me* that I have to have state-of-the-art facilities in order to better reach my students. *I kept looking for something* solid to make me want to "vote" for new facilities. It was left up to the reader to make the connections, which I feel is weak writing.

I agree with all the points about the grammar errors. Just about everything I had noted on the paper was covered by someone. I did note that on page 2 there were *two questions that would have been much stronger* if made into statements. One was "what's the need in rebuilding" in reference to the opposition's view. The other was "however is it fair to think education doesn't go past the teaching within classrooms?" I felt each questions [sic] was awkward and *would make clearer points if* turned into statement form.

DQ's narrative could be characterized as similar to the previous two in the sense that her original concerns emphasized a rather narrow range of Dimensions, and that over

time, they broadened to include additional Dimensions and a more balanced perspective. In her case, she moved from a fairly tight focus on evidence and support, which was viewed largely through the lens of correctly formatted citations that correspond to a correctly-formatted Works Cited page, to a wider range of concerns including complexity of argument, pedagogical approach, and reader response.

SS's Story. In the process of drafting this section, I became aware of how highly individualistic each participant in this study had become to me. Now as I read the posts in the discussion forum, I no longer have to glance at the color-coded letters to see who wrote each one. I can identify each person by the language and topic of their comments. SS initially became identifiable because of her focus on "attention grabbers," which was the first topic in several of her early posts:

> 1.1
>
> Wow! What a range of scores this paper received! It did have me pretty interested from the start. There was an attempt to grab the reader's attention, and I appreciated the sarcastic tone and anecdotal beginning.
>
> 1.4
>
> After reading the title, I felt that this essay had the potential of earning a 4.0. GO MSU! Then... I read what I thought should be the attention grabber. Is the attention grabber important? I emphasize that part of the paper quite a bit with my students, but it seems that most of these papers lack a starting sentence that really engages the reader.
>
> 1.5
>
> When I started reading, I thought, "Finally, an attempt at an attention grabber!" That is the one positive comment that I will post about this paper.

As in the previous narratives, SS's initial approach to writing was a fairly formulaic. The attention-grabber was merely the first item on a mental checklist that

included thesis, supporting paragraphs with topic sentences, and conclusion, and she was one of the main advocates for the five-paragraph essay in the discussion in Round Three. Her later posts, however, show a decreased attention to that form, and an increased emphasis on other elements. Her first points of concern in these posts from Round Five are focus, personality, points of the argument, and personal v. larger perspective.

> 4.5
>
> The paper's lack of focus made it very weak. I had a really hard time finding a specific thesis, and I found that the discussion of the video games, in addition to bored [sic] games, was too much.
>
> 5.3
>
> For me, the paper had the personality that the "Redesigning Recycling" paper was lacking, but it didn't have as strong of an argument. I'm not sure that it is possible to have the answers to such a big problem; however, the idea that the pharmaceutical companies should take back unused medications fell a little short of persuasion. What would be the incentive? Would this really be possible? Perhaps it is not the writer's responsibility to answer these questions, but it made the argument seem more hypothetical than realistic. Despite this, the paper was well-researched, well-written, and fulfilled the expectations of the assignment.
>
> 5.5
>
> This paper lacked the organization and support that a successful position paper should have. The personal perspective is okay, but the writer needs to also discuss the bigger issue and its implications.

As with the other participants, SS's posts in Round Five do not indicate that she lost interest in the attention-grabber, but that it became one smaller consideration in a larger universe of considerations. Here are some quotations from her Round Five posts, with italics pointing out the attention-grabber focus embedded in a larger whole:

5.1

-- Although *the attention grabber* is engaging, there seemes [sic] to be too many facts listed in the introduction, without transitions.

5.2

This was definitely one of the most well-written papers of the summer. The writer's voice was the only thing that I thought could be stronger. It was lost a bit in the formalities.

Postitives: [sic]

Great introduction

Solid thesis

Well-researched support

Strong counterargument

Complex sentence structures and advanced vocabulary

5.4

This paper was tough to score. Although it was *an interesting topic* that had much potential, the *cliche in the introduction*, the generalizations ("To most individuals..."), and the lack of sources throughout reduced the essay's credibility.

SS's comments, like the others, reveal a substantial shift during the study in the way she perceived student writing, and she began commenting in ways that were more reflective of the college analysis. Her online time and her posts were both shorter than the average of the participants, however, and by the end of the project she still produced a high ratio of summative to formative comments and less of the pedagogical approach than the more engaged teachers. Also, her comments in Rounds Four and Five were coded only once for Awareness of Writer and not at all for Reader Response.

OR's story. Of the six high school teachers, OR had the best academic background for teaching writing. She had a Masters Degree from the University of Michigan's English Education program, and had taken several courses that emphasized writing and writing instruction. From the beginning, OR's comments indicated a more complex, less formulaic approach to writing than the others. Her first post expressed a different perspective from the others' posts about the lack of a thesis statement in the introduction and shows a sophistication about the arrangement of the paper:

> I was also automatically looking for the thesis to be in the first paragraph. Looking back at the paper, I wonder if it wouldn't be best to leave the thesis at the end as the author has done, since her page one anecdote requires the reader to follow her reasoning before explicitly telling the reader what to think. Thoughts?

I think the last word, a question tossed out to other participants, also indicates that she had some experience in discussing writing with others and expected a dialog to ensue. Other participants mostly went about posting their own analyses of papers rather than engaging in dialog with others until controversies arose over the role of the five-paragraph essay and the role of conventions. It wasn't until Round Two that teachers began responding to each other's posts, posing questions for the group, and expressing both agreement and disagreement on a regular basis.

OR's comments throughout the project reflected an independent point of view. She often reflected about how her scores differed from others, and then proceeded with an analysis that also differed with others. She was fairly persuasive in her arguments, and others occasionally expressed agreement with her after reading her posts. This example is just the few first paragraphs of her post on Paper 4.4:

I found this paper hard to grade. I gave it at 1.5, and I see that I am way off the mark with everyone else. Like someone else mentioned, I do think this paper is competent. However, there are several major issues that made me give a lower score.

TONE. The author wavers between a sophisticated tone and a flippant, conversational one. The thesis (last sentence, first paragraph) is sophisticated, but the paper fails to deliver real evidence and theory to back it up. In a flippant way, the author gives hypothetical workplace and online learning scenarios that are speculative at best. I don't believe they serve as real evidence. Furthermore, the author makes comments that discredit his otherwise sophisticated tone:

-"[Online learner] has never really left the safety and comfort of his current environment, never expanded his social circle from that of his former high school, and never had to interact with people who have had entirely different life experiences from him." (Quite assumptive)

-"In these scenarios, there really aren't any fellow students to turn to for help, no helpful smart kid sitting next to you who can help you out" (Is the lack of a "smart kid" the issue at hand?)

-"However, this tends to not prove very effective as related to me by a co-worker who was in this exact situation" (= real evidence???)

-"In the end, I find myself curious as to what was going on in the mind(s) of those who felt that a purely online classroom was a wonderful idea." (sarcastic tone undermines the author's intelligence as he feigns ignorance)

The degree of difference from others OR displayed in her posts is reflected in her interrater reliability scores, which were the lowest of the group, reaching the satisfactory range of 0.8 or above only once, during Round Four. This did not seem to bother her at all, and in fact she posted far more extensive comments on the papers for which her assessment differed from the others, and her voice comes through in these posts as secure and unperturbed. Her discussion was invariably respectful and collegial, as was the case for all participants throughout the study, and she frequently expressed agreement with

others. Her participation in the study generally added both to the range and depth of topics and to the collegial quality of discussion.

OR may possibly have been in a position to provide leadership for the group and play a role in the formation of interpretative community. She had the professional training and skills of collegial discourse to do so. I looked at the printouts of the discussion forum for evidence that she might have been instrumental in the development of the group, but did not identify anything to support that idea. She often entered her posts later in the week than others, and her comments, while astute, were often a bit aloof rather than warm and supportive as some others were. She was moving during the summer, and during the project decided to give up teaching English for teaching Spanish, for which she was qualified and in which she would have far less paper-grading, so she was perhaps less emotionally involved in the project than other participants. At any rate, her participation did not appear to play a role more prominent than others in the formation of interpretative community.

Although the Table of Comments did not show any noticeable movement of OR's comments to mirror those of the college faculty more closely, her comments in the discussion forum do reflect an increasing pedagogical approach and awareness of the writer, and she "entered into the world of the student's discourse" on several occasions, commenting extensively on the various points and reasoning of the argument. Her independent opinions, supported by a more advanced knowledge of writing and writing pedagogy than the others, continued unabated, but even OR was influenced by the commentaries of the college group, particularly in regard to Complexity of Argument.

LL's story. As noted earlier, LL's participation in the project was limited by serious personal circumstances. Although she finished the project, her comments remained high-level summaries of one paragraph that did not reflect the depth of analysis and commentaries of the college teachers. She spent a total of a little more than 13 hours online. The only changes in her comments from beginning to end are: 1) increasing attention to complexity of argument, but this was minimal; and 2) increasing thoughts about revision (the Pedagogical Approach).

Summary of individual narratives. The review of the individual participants' comments was useful in describing how each person started with a different perspective on the teaching and assessment of writing and came to a different place – a shared place where complexity of argument, including counterargument and explicit expression of relationships rather than adherence to predetermined forms - was valued. They also developed in common a more pedagogical approach in their comments, including awareness of writer, of themselves as readers, and of the importance of feedback that balances positive and negative comments. The importance of the individual narratives is that they illustrate how each participant was able to adopt a larger perspective on student writing without wholly rejecting previous ideas.

Although I could not identify specific words or phrases in the discussion that directly explained the development of interpretive community, a picture materialized of gradual, nonlinear progress from a narrow set of ideas about student writing to a broader conception about how papers are best analyzed and described to advance student learning. Five of the six participants successfully transcended the limits of their prior conceptualizations of student writing to incorporate new perspectives, and they managed

to resolve the disparities between old and new ideas. The process of change seemed more organic than linear, which probably reflects the complexities of the learning process. In answer to the second research question about how change occurred, the best description is this process of gradual, incremental realization and acceptance of new ideas, while reframing and embedding old ideas to fit the new understandings. The other finding of note was that the participants who spent the most time on the project and commented on papers in the most detail appear to have made the greatest gains in approximating the perspectives of the college teachers.

Secondary Questions

This section relates to two sets of secondary research questions about the design and content of the Web site and the attitudes of the participants toward the project. The findings related to these secondary questions appear in the sections below.

Design of the Web site

Although I anticipated that design issues of the Web site such as navigation and graphics could hinder participation, these issues proved to be negligible. The basic design of the Web site was straightforward and pragmatic, and apparently it served its purpose well. In the interviews, only one person expressed even a minor concern, and that was about the login procedure required by the Angel course management system. Furthermore, only one of the participants suggested an improvement she would like to see in the initial version. She recommended using a split screen so that the student paper could be displayed on one pane, and the other pane could display either the discussion board or the PowerPoint slides. This would be a valuable improvement. If the opportunity

arises to engage a professional designer, better graphics could be designed to make the Web site more appealing, and a good designer could probably simplify the navigation, but in its present state it did not present any problems for the participants. The research questions about design of the Web site concerned the number of papers, amount of discussion, type of commentary, and so on, and they are addressed in a separate section for each question. A good generalization about these questions is that the project worked well as designed.

What volume of sample student papers is sufficient to produce enough learning for formation of interpretative community? Results of the scoring table (Appendix B, p. 166) show that a great improvement in interrater reliability occurred for five of the six participants between Round One and Round Two, but all of them made greater gains as the project continued. Only two of the teachers reached the desired rate of 0.8 before Round Four, but all of them had achieved it by Round Five. This indicates that teachers continued learning during the later weeks, so the project was not unnecessarily long, and a volume of 25 papers produced good results in formation of community.

Further, analysis of the discussion forum printouts indicated that movement of the high school teachers toward interpretive community with the college teachers occurred rather evenly over the course of the five Rounds. In most of the Dimensions and Perspectives where movement occurred, small signs of agreement began appearing early in the project, with increasing evidence emerging in subsequent Rounds, and continuing through Round Five. The only exceptions to this general pattern were the Awareness of Writer Perspective, which did not appear until Round Three and peaked in Round Five, and the Reader Response Perspective, where comments did not occur at all until Round

Four, and then became more frequent in Round Five. This finding also substantiates the conclusion that the volume of 25 papers spread over the five weeks was not unnecessarily long, as growth continued throughout the final week of the study.

What amount of and kind of analysis from college instructors is sufficient for formation of interpretative community? The amount and kind of commentary developed by the college faculty also worked well for this project, as evidenced by the pattern of increasing agreement by the high school teachers over the five Rounds, as noted in previous sections. The amount and kind of analysis is difficult to describe. This is true partly because there was such a wide range of requirements of the assignment and quality of papers, so the analysis of the college teachers naturally varied from one paper to another. Furthermore, since no rubric or other scoring mechanism was used to focus attention on particular features or qualities of writing, the college teachers were free to comment on whatever aspects of the writing they considered salient in each piece. The writing program at Blue Water College does emphasize the argumentative/persuasive essay with source material, and this provided a focus for much of the analysis, but the slides also showcase rhetorical features, stylistic issues, and problems with the conventions of written English. It is difficult to convey the depth and quality of analysis without using examples of the work itself, but the quotations of the college teachers' comments throughout the Findings section may provide a general guide to the range of comments and level of specificity with which the papers were analyzed. I believe the amount and kind of analysis developed for this project by the college teachers was the single most important element in the formation of interpretive community in this project, since it provided the model for the high school teachers to follow. As I noted in the

limitations section of the Methods chapter, a less capable and dedicated group of college teachers might not have provided content for the Web site that produced the level of formation of interpretive community seen in this project.

The finding that the amount and kind of analysis used in the study was suitable for formation of community was substantiated by the interviews. Answers to questions for the high school teachers about the slide shows in the midproject and final interviews indicated that they enjoyed viewing the slides and were eager to compare their own analysis of papers with the them. Most of the time they agreed with the college teachers and found that the slide shows confirmed their own views, but occasionally they were surprised by what the college teachers said, and in a small number of cases, they disagreed. They were universal in approving of the format, slide shows with embedded audio recordings. There was little variation among the responses of the high school teachers on this issue, and it is reasonable to conclude that the format, depth, detail, and quality of commentary from the college instructors used in this project worked well for the formation of interpretive community.

How much dialogue between college and high school teachers is needed to produce enough learning for formation of interpretative community? The college teachers participated in the discussion forum minimally, as I requested. I had asked them to monitor the discussion and contribute whenever clarifications or explanations were needed, but not to dominate the discussion. They complied with this request very well. At several points, in response to questions raised by the high school teachers, they commented at length on topics such as the relative value of the conventions of written English (spelling, punctuation, and so on) compared with higher-order concerns such as

logic and rhetoric. A few times they responded with short answers to queries, but most of the time the college teachers simply observed the forums without comment. Judging from the analysis of the discussion forums, this was a successful balance. The high school teachers quickly followed the suggestion of one of the college faculty to require revisions in the system of grading. The high school teachers asked questions of the college teachers on five occasions, two of which generated some back-and-forth discussion of the issues. Most of the time, however, the high school teachers addressed their comments to each other, and their comments provide the great bulk of the discussion forums. Interviews with the high school teachers indicated that the amount of interaction with the college teachers was satisfactory. With the exception of one teacher who wanted an answer from the college faculty about the five-paragraph essay, the teachers agreed that the BWC participation was "about right." None of the high school teachers was critical of the college teachers' comments, although they maintained some individual opinions about the value of the five-paragraph essay form and the role of the conventions of written English. I have concluded then, that the amount and type of discussion between high school and college teachers generated in this project, including occasional in-depth responses and short answers to inquiries by college teachers, but generally allowing the discussion to remain in the purview of the high school teachers, was conducive to the formation of interpretive community.

How much practice in scoring do teachers require for formation of interpretative community? The results of the quantitative analysis showed that interrater reliability reached a peak in Round Four (0.94, compared with 0.88 in Round Five), so if interrater reliability were the sole goal of the project, a quantity of 20 papers would have been

sufficient. The qualitative evidence suggests that the teachers' analysis of the Dimensions and Perspectives continued to evolve well into Round Five, however, so if results beyond interrater reliability are desired, at least twenty-five papers are needed. The question of how much learning would occur after 30 or more papers remains for another study, but for the purposes of this study, the finding is that 20 papers for the purpose of high interrater reliability, and 25 papers for the purpose of greater approximation of the college teachers' analysis of papers, provided a good amount of scoring practice.

How much dialogue among high school teachers is sufficient for formation of interpretative community? Interview responses at the conclusion of the study indicated that most of the participants felt that online dialog with each other was the most meaningful and learning-producing part of the project for them. They felt the slides with the college faculty's analysis were important, but the discussion with their peers was mentioned first by all six teachers as the part of the study they found most meaningful.

As with the previous questions in this section, evidence from the discussion forums suggests that the amount of dialog generated among the high school teachers in this project was sufficient for the development of interpretive community. The teachers did not interact with each other much during Round One, during which they posted their own analysis of the papers and did not respond to each others' posts. It was only after issues arose that were of concern to most of them (the five-paragraph essay and the relative value of the conventions of written English), that responses to other posts became a common practice and any sense of community emerged. In the case of this study with these particular participants, the amount of discussion generated by the 25 papers can be considered sufficient for the development of interpretive community.

Summary. The findings for this set of questions, then, are that a number of 25 papers, spread over five weeks, with roughly the same amount of commentary from the college faculty and discussion among the high school teachers as were generated during this project, is at least minimally appropriate and useful for the development of interpretive community. Future trials of this type of online community could experiment with different numbers of papers, discussion requirements, and expert commentary, but the results indicate that the design of this study was practicable, and participants in this study were generally pleased with LINC as currently designed.

It should be noted that, within the parameters of the design in this study, there was great variation among the high school teachers' participation, as noted in the section on individual narratives, and that greater participation yielded greater gains in approximating the college teachers' analysis of papers.

Attitudes of the Participants

All ten participants expressed favorable attitudes about the project in their final interviews. The views of the high school teachers and college teachers are reported separately here, and since the quotations are much shorter in this section, they are embedded in the paragraphs following APA style.

High school teachers. The type of response that stands out in all the interviews with high school teachers is that they enjoyed, and profited from, the opportunity to see how other teachers viewed student writing. All six teachers voiced this opinion about the study in one way or another. Several people commented on the value of learning that other people shared their views. The word "reinforcing" was used by two teachers in this

regard, and another called it "validating." An equal emphasis was placed on the experience of seeing opinions that differed from their own. The words "challenges" and "disputes" were used to describe that part of the experience, and one teacher said, "It kept us on our toes." Another said the project helped to "clarify weaknesses, biases, and approaches." These comments from the interviews coincide with the coding of the discussion forum, which showed an increasing number of Expressions of Agreement as the study went on, but also several codings of Expressions of Disagreement in Rounds Two, Four, and Five. At no point in the interviews or discussion forums did anyone express negative feelings about disagreement, such as frustration, anger, negatively viewing another person, and so on. Rather, the usual response when a participant initially disagreed with someone else was either to reflect and to then agree, or to express an introspective reaction such as, "I have learned that I have to accept the fact [that students don't know mechanics] and to be more realistic in grading."

In response to the question about what they did not like about the study, there were several answers, but each one was an individual response that was not echoed by the others. One teacher was not good at keyboarding and the discussion forum was painfully slow; another one did not have high-speed Internet service at home and had to travel to her school to do the project. At one point, another person said she "felt off" compared with the others, but on checking back found that her comments were often supported by others. Another teacher, who described herself as the "early bird," would often post early in the week and have to wait a long time to see how others responded. She was "anxious to hear others' reactions." Another was still confused about the BWC stance on the five-paragraph essay, and wanted clarification. One mentioned that having the Works Cited

page and knowing the exact assignment for all the papers would have been helpful. One person said a sheet explaining exactly how to post on the forum would be good. One person found that writing her comments out in a Word document and then copying and pasting them into the discussion forum saved time. One person had nothing negative to report.

Four other questions elicited uniformly positive replies. Asked if participation in the study would likely affect their teaching in the future, all six said yes, although the changes they thought about varied. One would teach development of a single paragraph (which may run counter to current thinking on writing pedagogy); another would give more direct instruction in mechanics; one plans to change her rubric; and one said merely, "More resolve." Two said they would begin to use peer editing and classroom editing. Second, a question asked if the project had been worthwhile for them, to which all six responded with a yes, some emphatically. Third, they all responded positively to the anonymous, asynchronous online experience that afforded them great flexibility of schedule and the opportunity to work at home. And finally, they were asked if they would recommend the project to other teachers if it were continued, and they all answered yes. One had already told the "history department" at her school about it and suggested they try class and peer editing.

All six high school teachers also expressed in some way or another feelings of respect for and valuing each other. One teacher referred to the group as "kindred spirits" and said they were not "a typical group of teachers." Two teachers said they thought the project should continue in the future as a "refresher," and three of them asked if they could use the PowerPoint slides in their own classrooms. (They could not, since I did not

have permission from the student authors of the papers to use them in that manner, but it is something I would like to prepare and offer sometime in the future.) Finally, two of the teachers suggested that they would like to meet the others, and the others agreed. We set a time and place for a gathering. Two high school teachers who were in the process of moving their households did not attend, but the other four did. Only one of the BWC attended; the others were out of town. (I believe they had delayed vacations until the project ended.) It was an enjoyable get-together. As each new person walked up to the table, the others had fun guessing who it was. The conversation was lively and centered on teaching. Although this type of social interaction and personal engagement is not described as an essential element of learning or interpretative communities, the meeting did underscore the belief that the project had meaning and value for these teachers.

Considering that teachers of English generally feel burdened by the extensive paper grading they must do (Belanoff, 1991; Huot, 2002), and that the six participants had voluntarily graded 25 papers and commented at length on them during their summer vacation time, I think their favorable attitudes are a highly positive outcome of the project.

The college teachers. The four college participants were also interviewed for about an hour each at the end of the project. Unlike the high school teachers, their work in the project was hindered by technical issues, which are summarized in Appendix L, p. 189. Also, unlike the high school teachers, the college teachers were required to collaborate, deciding in pairs how to score a paper and what strengths and weaknesses to point out, taking into account input from the other two teachers. They did not always arrive at a single score, and on a couple papers, their evaluations were far apart.

Furthermore, the idea of collaboration was not natural for one of the members, who continued writing out his own commentary for each paper and expected to record it in whole at each session, and then pass the mike to his partner. Some of this problem was overcome with editing, so that his comments were interspersed with his partner's on each point. It caused the editing process to be extremely slow, however, and his partners preferred the more collaborative style they developed with each other, in which they discussed each paper at length before recording and jointly developed an outline of points to include in the commentary. The college teachers also agreed with the high school teachers that not having the exact assignment and the Works Cited page was problematic. In the only other negative view expressed about the project, they agreed it was a lot of work.

This group had several positive views in common with the high school teachers, however. They all enjoyed the collegial dialog and hearing each other's perspectives on the papers, and they all thought it would be beneficial for their own teaching. In addition, two of them thought LINC would be valuable in teacher education programs as well as for faculty development, especially for new teachers. In response to the question asking if, knowing what they know now, they would still have accepted the invitation to participate, all of them answered yes. Again, as with the high school teachers, considering that people gave many hours of their summer vacation time to grading and analyzing student papers, their responses indicate a very positive attitude.

Summary of Findings

The first of the research questions, whether interpretive community could form in the type of online environment constructed for this study, can be answered affirmatively, with strong evidence. First, high school teachers achieved high correlations of interrater reliability with college teachers in the scoring of student writing. Second, qualitative evidence from the discussion forums indicated that, by the end of the project, high school teachers approximated the ways the college faculty analyzed student writing in seven key respects: complexity of argument, counterargument and concession, explicit connection of ideas, pedagogical approach, awareness of writer, reader response, and positive focus. These developments comprise strong evidence that an interpretive community did form over the course of the five weeks of the project. What the results of the study did not achieve is an explanation of how that community developed in an immediate, direct way that linked specific comments or experiences with changes in analysis of the papers. The individual narratives that describe each teacher's initial thinking about student writing in comparison with a more comprehensive, balanced perspective at the end of the study illustrate how gradual, incremental change occurs in a nonlinear, nonrational fashion, and how old ideas are incorporated into new ones in the process of learning in a community.

The data also suggest that the general design of the online project (25 papers spread over five weeks, requirements to score papers before knowing how others scored them, discussion forums dominated by high school teachers but occasionally engaged in by college teachers, and the PowerPoint slides with embedded audio recordings of the college teachers' analysis of papers) provided enough discussion, practice, and expert commentary for the development of interpretive community. Further, responses by both

the high school and college teachers indicated that they believed the project was a good investment of their time and energy, that they liked the online format, and that they enjoyed and profited from the collegial dialog with their peers.

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

A brief summary of the findings of this study in relation to the research questions is offered here to provide a foundation for the implications and recommendations discussed in this chapter:

1) A meaningful degree of interpretative community formed among participants.

2) Community formed through a process of gradual, incremental realization and acceptance of new ideas while reframing and embedding old ideas to fit the new understandings.

3) Participants who spent the most time on the project and commented on papers in the most detail made the greatest gains in approximating the perspectives of the college teachers.

4) The design of the Web site proved to be adequate for formation of interpretive community and created no problems for the participants.

5) Attitudes of both the high school and college teachers were highly favorable toward the project.

The remainder of this chapter is organized into three sections: First, I offer some general thoughts about *how* community developed, by relating the individual narratives to a fuller explanation of the theories of communities of practice and collaborative learning. Second, I suggest some questions for future researchers to explore, including applications of the idea of assessment communities that, if found workable, could help solve some other problems in education. Third, I provide some implications for college

administrators and policymakers as they make decisions about writing assessment and faculty development in the current political and educational environment.

Conclusions about Community and Assessment

First, both the quantitative and qualitative evidence supports the conclusion that a meaningful degree of interpretative community was formed in this study. At the end of the study, the high school teachers as a group analyzed and commented upon student writing in a manner far more like the college teachers, and far more consistently with each other, than they had at the beginning. This finding was obtained not only for the particular features of student writing that participants commented upon, but in their general approach and perspective about the assessment of student writing.

I believe the importance of these findings lies in the quantitative evidence and qualitative evidence considered together. Although improvements in interrater reliability over the course of the study were substantial, similar results are obtained by testing companies when they norm scorers to rate essay tests, so they are no great surprise. What is meaningful in this case is that the online assessment community, by its very nature, should make high levels of interrater reliability difficult to achieve, for three reasons.

First, the range of assignments and the depth and complexity of student work in this study more closely replicate real-life writing than do standardized test prompts. The tasks assessed in this project mirror the actual tasks given to students in first-year composition courses at BWC, because that's exactly what they were (in 22 of the 25 papers – three were from other institutions). The practice of using actual samples of student work produced in coursework affords this kind of assessment near-perfect

construct validity: there is an almost exact match between work assigned and work assessed. This contrasts with the kind of writing tasks, or prompts, used in standardized tests, which are carefully developed and selected to narrow the range of possible students' responses so they can be scored quickly by minimally-trained scorers (Huot, 2002; Yancey & Weiser, 1997). Using the diverse range of samples of student work collected in college courses establishes very high validity for this method, but it could be expected to result in low reliability. Therefore, the achievement of high levels of interrater reliability is especially important.

A second reason that the findings of interrater reliability in this study are remarkable is related to the qualifications of scorers. It might be assumed that training scorers to reach a high level of interrater reliability would be more difficult if they lack college teaching experience and advanced degrees in English or related fields, and that it would be easier to train better qualified people. The opposite is generally true: professionals with advanced degrees and teaching experience generally bring more complex and divergent perspectives to the task of analyzing student work, and their opinions are well established (Broad, 2003; Huot, 2002; Elliot, 2003). They "see" more things in student writing, and they find it harder to ignore what they see. The practice in the commercial testing industry of hiring part-time, often seasonal scorers, who usually hold only Bachelor degrees (not necessarily in a relevant field), with no benefits or possibility of advancement, at the rate of $12 - $15 an hour, is what makes the writing test industry profitable. Such scorers are easily trained to produce reliable scores, as long as the scoring rubric and writing prompts are carefully controlled. The fact that ten participants with Master's degrees in English or related fields and with many years of

teaching experience were able to reach a high level of interrater reliability through participation in the Web site is noteworthy.

A third reason that the high level of interrater reliability is impressive is that the study did not use a scoring rubric. Scoring rubrics used by testing companies are tailored narrowly to the particular prompt, and they typically ignore many of the features of good writing that first-year composition courses emphasize, such as the evaluation and use of source material and rhetorical strategies and devices. When scorers are trained to high degrees of reliability using a specific rubric, they read narrowly from a single viewpoint (Broad, 2003), and the constricted viewpoint of the scoring process reduces validity (Huot, 2002). In this study, scorers were unimpeded by a rubric or other set of values imposed from an authority outside the academic context to which it is targeted, so that the values underlying the assessment decisions in the study were in fact those of four of the instructors who actually teach the courses in which many of the region's high school students will enroll in the future. This is another feature of the online assessment community that establishes high validity, but might be expected to produce low reliability.

The high construct validity of this exercise, therefore, is what makes the reliability numbers meaningful. It is one thing to obtain high interrater reliability on tests in which scorers are trained to produce reliable scores using a rubric imposed by test developers that may or may not closely reflect the values of the faculty (Huot, 2002). It is more difficult to obtain high interrater reliability with faculty who hold diverse degrees and opinions – and who cannot be fired immediately if they fail to score student work according to pre-established norms. It is only when assessments very closely match the

actual work of writing and evaluation that results in interrater reliability become meaningful. It is especially true that assessments become useful as guides to teaching practice only when they maintain high construct validity. The key to meaningful and useful assessments that have some potential to actually improve instruction, then, lies first in validity, and only secondarily in reliability.

But validity based on what construct? If the construct of writing being measured had been derived from the understanding of the high school teachers in Round One, it would almost certainly proceed from the form of the five-paragraph essay: an introduction with a thesis statement (usually the last sentence in the paragraph), at least three supporting paragraphs beginning with a topic sentence that relates the paragraph to the thesis, and a conclusion that summarizes the points and restates the thesis in different words. An "attention grabber," solid mechanics, and good use of evidence would be second-level concerns, and the use of apt words, complex sentences, literary devices, and rhetorical strategies would slide into third place. What would be missing is the point-by-point scrutiny of source information and the student's own reasoning and assertions, which are at the heart of critical thinking. If the construct of good writing were developed from Round One thinking, in other words, it would look much like K-12 writing tests administered by states for the purposes of accountability and like the SAT and ACT writing tests used in college admissions. RR was essentially correct in his description of how scorers of writing tests work: check for certain elements of the argument quickly and then make an overall judgment about voice, conventions, style, and rhetoric. There is no time for studious examination, fact-checking, or reflection, and there is no opportunity for students to locate, evaluate, and incorporate source material. So, although SAT and

ACT writing tests may be generally a notch higher than the teachers' Round One analysis, the distance is not far; but it is a great stretch from there to the detailed investigation into specifics of the argument modeled by the college teachers and reflected by the high school teachers in Round Five.

I would like to emphasize certain qualities of the Round Five discussions. The teachers had, as I described it earlier, entered into the universe of the student's discourse. It was as if they became engrossed in the argument themselves, sometimes suggesting to the student better moves to make, sometimes playing devil's advocate, sometimes thinking aloud. They retained the authority to make judgments about the paper, to say what worked and what didn't, but they did so more as a partner or guide in the writing adventure than as an impartial and distant judge in residence outside the process. In the context of this project, teachers could not actually talk to any of the student writers, but among their Round Five comments are frequent glimpses of what they would say to the student and how they would frame their responses. More than any of the data I collected and searched, I think this sense of dialoging with students about their topics - and I would go so far as to say - building community with students, explains the *how* of research question #2 (*How did interpretative community, if any, form?)* that was difficult to answer with the empirical evidence of specific comments. And - I will offer a conjecture here – it is possible that the community teachers were preparing to extend to students in Round Five originated among the college teachers as they shared notes, developed outlines for the recording sessions, and arrived at consensus on scores. The college teachers developed some form of community among themselves, which is apparent in the quality of their recordings and the interviews, and then extended the community to the

high school teachers in the form of PowerPoint slides with embedded audio and comments posted in the discussion forum. The high school teachers moved a considerable degree from the peripheral position of novice in Round One to a position closer to full professional status in Round Five, from which they could then begin introducing their students into the community.

It would be appropriate to note here that, just as college teachers were privileged over the position of the high school teachers in the design of the study, this idea of high school teachers extending their interpretive community to students likewise privileges the knowledge of teachers over that of their students. The privilege of "experts" over "novices" is inherent in the literature of learning communities and communities of practice, however, and does not dislocate the theoretical framework of the study from its foundation in constructivism. The knowledge of the expert group is not founded upon a special access to absolute truth unavailable to others, but upon mutual engagement and shared discourse with each other over time that created that knowledge in the first place, and it is understood that successive generations of practitioners will continue to modify and refine the knowledge. In this sense, assessment communities may be understood as open systems whose boundaries are responsive to the environment so that they are constantly influenced by new ideas, resources, and energy (Morgan, 1997). It is quite likely that, although not a subject of inquiry in this study, the college faculty's thinking about writing was influenced in some ways by the high school teachers. High school students, in turn, are likely to influence their teachers. Assessment communities operating on a continuing basis, then, embody knowledge in the manner of a living organism that constantly interacts with its environment, taking intellectual nourishment and energy

from it, and modifying its internal structure to maintain homeostasis. This view contrasts with a positivist view in which knowledge is static and isolated, embodied in textbooks, and untouched by the learners who acquire it.

Because the idea of professional community underlies much of my interpretation of the findings, I would like to review some points of Etienne Wenger's (1998) book, *Communities of Practice: Learning, Meaning, and Identity,* that are especially relevant to this study. I believe they are relevant not only because they describe results of this study, but because they illuminate possibilities for future applications of assessment communities. Wenger says that *practice* is a social activity in which participants negotiate both explicit and tacit meanings in historical and social contexts. Learning occurs in a community of practice through mutual engagement, joint enterprise, and shared repertoire. Assessment understood through Wenger's lens, then, would be a social practice that produces criteria, both explicit and tacit, specific to the social and historical context in which it is exists. The regional flavors of BWC and the students shared by both high school and college teachers did not appear in high relief in this study: their words sound as if they might have been written by any group of high school and college writing faculty. If the writings and analysis were compared with those of other regions and contexts, an urban college or an East Coast school, for example, differences would probably become apparent. That assessment criteria are local and bounded by social contexts, however, is not a fault of an assessment community, but the source of its validity.

Further, Wenger says, mutual engagement among the participants would be organized around the work they do. In this case the work occurs online, where the

discussion forum is the location of interactions among all participants, who are focused on the common purpose of assessing student work. The joint enterprise of such work would "reflect the full complexity of mutual engagement" (p. 77), which is echoed in the great variety of comments about each paper and the commenter's interactions with each other. Individual voices remained distinct throughout the LINC project, and there is a clear sense in the discussion forums that each participant, while considering and sometimes adopting the views of others, retained a position of individual authority and judgment, while at the same time finding ways to show respect and consideration for each other. Another point in Wenger's explication of community is that the joint enterprise would "create among participants relations of mutual accountability that become an integral part of the practice" (p. 78), a phrase that describes comments of the teachers expressing both agreement with, and challenges to, each other. The "repertoire of knowledge" shared by members of Wenger's communities reflects the history of negotiation within the community, but remains ambiguous and open to future negotiation. This quality too, was displayed in the discussion forums. There are no instances of college faculty explaining, lecturing, or rule-setting in a positivist sense, even in response to a teacher's questions about the role of conventions in grading. Rather, the faculty commentaries display many points of agreement, and several instances of disagreement, in a respectful and collegial manner. Their analysis and scoring decisions are presented less as final conclusions than as mutual agreements, which allow for the possibility of different perspectives. Their verdicts are often phrased in terms such as, "We thought...," "One thing we liked...," or "Where the writer seems to have missed...," language that is equivocal enough to be open to other interpretations and perspectives. Finally, Wenger

says, learning in a community of practice would not be a special category of performance, but an embedded and ongoing process of the practice. This too describes the record of interaction among the teachers in this study. Aside from a few comments coded as Introspective Comments, the learning occurs not as a separate feature of participation, but as an ongoing process embedded in the interaction.

I would add to Wenger's concepts of practice some dimensions of Mintzberg's (2000) notions about professional practice. Speaking of surgeons, teachers, and social workers, he said that standardization of practice ensures that "everyone knows what's going on" (p. 51), but at the same time, the complexity of the required skills "ensures that considerable discretion remains in their application" (p. 51). No two practitioners practice exactly the same way. They utilize their professional knowledge, but must apply it with judgment to individual cases that often do not present the clear-cut archetypes of indicators found in textbooks. The practice is complex and difficult to measure. Following the discussion forums and PowerPoint slides in this study has elucidated for me, once again, how very difficult and complex assessment of writing is. Merely describing the features and qualities of writing, labeling them with accurate terms, and classifying them in clear and mutually exclusive categories was a daunting task, and the more earnestly I sought guidance from college curricula and scoring rubrics from various institutions, the more confusing the task became. Even after I had developed categories and dimensions that seemed to capture the data suitably, it was impossible to classify every one of the multitude of features presented in the 25 student papers with perfect certainty. It required judgment. In short, in both Wenger's and Mintzberg's terms, the activities teachers engaged in during this study can be described as professional practice.

If assessment of writing were elevated to the status of professional practice in teacher preparation and development, and if it were accorded the same stature as diagnosis in medical education, it might become an important step in improving student writing.

Contributions to the literature by Wenger on communities of practice, by Mintzberg on professional practice, and by Fish on interpretive communities all help explain what happened in this study. Viewed as a professional practice, assessment is not a matter of measuring data by specific criteria and noting which tick marks are reached; it is the application of professional knowledge to individual cases that present some degree of ambiguity, contradiction, or incompleteness, and therefore require judgment. Assessment is not the application of external, absolute criteria to isolated cases, but the application of negotiated, contextual, and ever-changing meanings to artifacts produced within the local context. Most importantly, the stability of meanings (or reliability of results) is produced by common values and shared discourses within the community of practice. The greater degree of accord among high school and college teachers in Round Five is inferred to be not so much the result of the acquisition of new knowledge (although that may have played a part), as it is the result of participation in a community of practice.

Kenneth Bruffee's (1999) book on collaborative learning is also relevant to understanding the processes of change that occurred in this study. He explains that the process of learning is not acquisition of cognitive forms and frameworks, but a social process beginning at birth, in which the infant is inducted into membership in the mother-child community constituted by looks, touches, and gestures. The growing child does not replace the community with mother, but instead,

> ...subsumes membership in the closed community of infant and mother, nesting it more and more deeply within larger, more comprehensive, and ultimately more rewarding communities. From the very beginning, *knowing* [italics original] is a complex of such memberships in infinitely nested, overlapping linguistic communities. Learning is reacculturation from one community to another" (p. 188).

The nesting of communities occurs much as Russian dolls of various sizes nest within each other, or measuring cups nest in the cupboard. Bruffee illuminates the process of change observed among the teachers in this way: first, he makes clear that "knowledge" as he speaks of it is socially constructed, and that learning occurs by induction into new communities. Much of the learning is linguistic. In this light, changes within the high school teachers can be understood as a gradual adoption of the language, values, and perspectives of the college teachers. For example, RR continues his detailed analysis of grammatical and stylistic forms in student writing, but subsumes them into the Pedagogical Approach and Positive Focus manner of the college teachers, and he pays far more attention to the particulars of the argument. Similarly, SS is still looking for attention-grabbers at the end of the study, but begins using the term, "engaging," which reflects the language of the college faculty. FR has not forsaken her knowledge of the Six Traits, but has incorporated them into a more complex and detailed view of writing. All the teachers, including the one who defended the five-paragraph form to the end, shifted their understanding of coherence from a function of the five- paragraph form to a consideration of how relationships and connections are expressed. Bruffee's vision of learning goes beyond this kind of "additive" process, however, and insists that learning of higher cognitive skills is not a strictly cognitive process, but primarily a social one. Even in medical education, he points out, students who learn to diagnose patients in a collaborative manner become more proficient than those who learn in a cognitive model.

A curious result of the study is that the data from interviews with the high school teachers are largely lacking in any sense of awareness of how much change had occurred in their perceptions and analysis of student writing over the course of the project. The evidence of change is apparent in the discussion forums, yet when asked to describe any changes they observed in their own scoring or analysis of papers, all six made only the most general comments and could not name specifics. In fact, three of the six indicated that they believed their analysis and scoring of papers was similar to the college teachers' evaluations throughout the project, rather than moving closer during the study. None of my research questions was directly addressed to the level of cognizance participants displayed about their own learning, so I did not attempt to understand this phenomenon, but it would be an interesting question for future research.

At this point it may be appropriate to ask: is this discussion about how teachers assess student writing, or about how teachers *learn* to assess student writing? I would answer, both, because it is essentially the same process. Teachers assess their students' essays by writing comments on papers, conferencing, and guiding class editing and revision lessons. They are able to give feedback to students only insofar as they can "see," in Fish's sense, the qualities and features of writing that are desirable in student essays. To raise the teaching of writing a few rungs up the ladder, to a level of complexity and pedagogical approach appropriate to college-level writing, it is first necessary for teachers to gain full sight of that level. The findings of this study support the notion that the opening of teacher vision to higher-level dimensions of writing can be accomplished to a considerable degree in community, which may provide the basis for some changes in their practice of teaching. Whether they are called learning

communities, online collaborative communities, interpretative communities, or assessment communities, they can be conceptualized as locations where teachers work from novice status toward expert status in a professional practice by reacculturating to a new community, and then have the opportunity to turn back again, outward toward their students, and begin to introduce them into the community. Because I believe that assessment is a special case of interpretative or learning communities that has great potential for solving the thorny problems of improvement in learning outcomes and accountability, I will refer to them hereafter as "assessment communities."

Closing the Loop with Assessment Communities

Another way to look at the findings of this study is to think about that highly-prized aim of educational research, "Closing the Loop." In many depictions, a circle represents a cycle in which curriculum is designed in Step 1; curriculum is implemented in Step 2; and learning is assessed in Step 3. Step 4 is problematic, because it is supposed to utilize the results of assessment to improve curriculum and start the cycle over again (Palomba & Banta, 1999). In reality, however, the loop is often broken at Step 4, and assessment findings somehow do not manage to complete the curve back to Step 1. Various strategies for "closing the loop" are popular topics at conferences and in journals. It has been suggested that the failure at Step 4, at least in the pedagogy of writing, is a result of a divorce between instruction and assessment, between teachers and testers, that occurred with the development of multiple choice tests at the beginning of World War II (Elliot, 2005, White, 1994). Faculty complete Steps 1 and 2, but educational measurement specialists complete Step 3, and when the results are communicated back to

faculty, they make little sense. Writing and liberal arts faculty work from different paradigms and assumptions than measurement specialists. They read different literature, attend different conferences, think in different terms, and speak a different language. In short, they belong to different communities. The results of standardized tests, especially, bear little resemblance to the ways faculty think about curriculum, student learning, and teaching challenges; and faculty are not able to translate the psychometric language of test results into meaningful and usable terms. It could be said that the brains of faculty are filled with Aristotle, Shakespeare, Newton, Jefferson, philosophy, and literature. The brains of testmakers are filled with Cronbach's alphas, Pearson chi-squares, standard deviations, covariance, and regressions. When the testmakers hand the results of their tests to the faculty, the faculty hope to find a key to learning, but instead find bar graphs. Knowing what percentage of students achieved at what level is not helpful. It does not inform faculty about how to move students upward on the achievement scale. Assessments conducted in this study, by contrast, are grounded in the work, language, and perspectives that fill teachers' lives, and so have meaning and resonance. High reliability in this context is a substantial achievement.

A Web-based online assessment community such as LINC offers one means of repositioning the assessment of writing back into the purview of faculty, so that the features and traits of writing, the terminology, and the ways of thinking used in assessments are the same as those of curriculum and instruction, and they are embodied in the same group of people. Although this study makes no claims that teacher learning achieved during the project will necessarily result in changes in teaching practice, the method does hold some potential for causing the Loop to curve back to Step 1, naturally,

of its own accord, because the teachers and the assessors would be the same body of people, belonging to the same interpretative community. The meanings would be stable, and people would "see things" the same way.

Concluding Thoughts

The notion of learning communities has been linked to faculty development in literature for some time now (Bruffee, 1999), and the results of this study amply support the valuable connection between the two. Just as students learn best in socially-mediated contexts that provide for exchange of ideas, challenges to thinking, and generation of new ideas, so faculty learn best in transition groups that offer both support and challenge, confirmation of old ideas, and expansion of new ones.

What is new in this study is the injection of assessment to the mix. Although a conception of assessment as a professional practice embedded in community is dramatically different from the conception of assessment emanating from commercial testing companies, departments of educational measurement, and accrediting agencies, it holds some promise for solving difficult problems in education. Assessment immersed in professional community, conducted by groups of people who employ their common interpretations of reality even as they evolve new ones, and bolstered and sharpened by collegial dialog, holds some promise for assessments that are far more valuable to faculty and more likely to improve instruction than the standardized tests that are frequently recommended now (Spellings Commission, 2006; Carey, 2009; Lederman, 2009).

In contrast with standardized essay tests, assessments like the one in this study are conducted on a broader range of student work, and they are sensitive to a far greater

range of complexity, detail, and nuance. They afford faculty a deeper view into the capabilities and limitations of their students, so they provide more usable information about student performance for the purpose of improving instruction. For the purpose of research on teaching, the transcripts could also be used as a rich source of information about how faculty read student papers. The greatest impediment to the adoption of assessments like LINC may not be lack of knowledge about their superiority over tests as a platform of understanding from which to improve teaching practice, however. It may be instead concerns about the resources of time, money, and intellectual energy required for their implementation. The method studied here does require investment of resources.

First, the method requires a significant amount of time. Two of the four college faculty probably devoted far more than the 20 hours they committed to during the recruiting process. One of them took a personal interest in the problem of transmitting large audiofiles and other technical issues, and another assumed the role of organizer and problem-solver. Both of those roles were necessitated by the startup process and might be avoided in later projects, but it was clear that a fully-engaged, onsite manager and some technical support is essential, in addition to the work of creating the expert analysis and commentary,. Also, at least two of the high school participants, the ones who demonstrated the most change over time, spent far more time than the recruiting materials called for. In calculating the amount of time faculty participants are asked to invest in this kind of project, it is important to acknowledge the time spent offline in reading, analyzing, and preparing notes for the discussion forum.

In addition to the precious resource of time, an institution would need to make a financial commitment to the project. The Web site itself can be quite simple technically,

but later iterations of it would no doubt profit from the expertise of professional designers and programmers. The amount of Web space required for the project would have to be expanded as the number of participants expanded. Finally, the development and administration of this type of assessment would predictably be more complicated than signing a contract with a commercial testing company and meeting their needs for scheduling, access to students, motivating them to do their best, and dissemination of results.

Some of the demands this method makes on faculty and administrators may be balanced by gains in the validity of the assessment, usefulness of the results, and meaningfulness to both students and faculty, but other incentives would probably be needed to recruit large numbers of faculty. Inducements for faculty participation could include recognition of achievement of a specified level of interrater reliability or time logged online, for the purpose of promotion and tenure, or, in the case of adjuncts, consideration for tenure-track positions. In any system where faculty development is rewarded, participation in an online assessment community would be appropriate.

At any K-12 level, credit toward continuing certification, which is required in most states, would be an attractive incentive for many teachers. At the secondary level, priority in assignment to teach upper-level and AP courses might be an attractive incentive for veteran teachers who no longer need continuing education credits. Another possibility is that local or regional charitable foundations would provide honorariums for participating faculty. At the level of state government, a department of education could issue certifications to individual teachers, schools, or school districts for specified levels of participation or achievement in online assessment communities. There might be

several levels of certification: first for initial participation; then for taking leadership roles in the continuing process of refining assessments at the local level; and eventually for managing the state-wide process of updating and managing the process of assessment of writing at every level.

Without question, shifting writing assessment from commercial companies to writing faculty would entail management issues and costs. The costs might be counterbalanced, however, not only against the costs of standardized testing, which are considerable, but against potential gains in faculty development, which might also be considerable. If further research on online assessments communities substantiates the degree of change in teacher scoring, analysis, and commentary in evidence in this study, and especially if gains are made in follow-up studies of student writing, the cost of this assessment might compare favorably with current expenditures for assessment and faculty development. In other words, online assessment communities can potentially kill two birds with one stone, so the cost must be estimated in the light of total benefits.

An important component of this study, and potentially an incentive to teachers, was that the entire enterprise (except for the interviews which were necessitated by research needs rather than the assessment itself) was conducted online, allowing participants who had high-speed Internet connections at home almost unlimited freedom in their schedules and the enjoyment of the comforts of home. The fact that both high rates of interrater reliability and convincing qualitative evidence of the formation of interpretative community were achieved in this medium has vast implications for the use of online assessment communities to link faculty across every kind of geographic, temporal, cultural, and institutional divide, and to do so at very low costs.

Before concluding this chapter, I must acknowledge that, although the online portion of this study is new, the essential idea of engaging teachers in communal conversations about assessment of student work is not new at all. It happens in one version or another in many colleges, universities, and school districts on a regular basis, and it has a history going back at least as far as the College Board entrance exams of 1900. What is new about the online version is that, by shifting the dialog from face-to-face meetings to the Internet, it becomes an activity that faculty can engage in without scheduled meetings, which helps make it a more acceptable demand on their limited time, and, in spite of the necessity of typing one's ideas out instead of merely speaking them, it makes the project more workable in busy faculty lives. In fact, the act of typing may provide a second or two for self-editing that produces more thoughtful and carefully considered responses. The online assessment community also holds forth some promise that both high-quality faculty development and valid, reliable measures of student writing ability can be achieved, and if brought to scale with large groups of teachers, with minimum expense.

As a final thought, I would like to introduce yet another acronym to the educational lexicon: WYTIWYG. Pronounced "whitiwig," it stands for What You Test Is What You Get. The educational enterprise is a behemoth that turns only slowly, but it does turn, and eventually what we program into the system by way of assessments will come chugging out the other end in the form of outcomes. The landmark study of curriculum by Diane Resnick and David Resnick (Huot, 2002; Yancey & Weiser, 1997) found that whatever is *not* tested tends to disappear from the curriculum, and the current focus on preparing students for 20-minute writing tests in K-12 education is evidence that

whatever is tested will eventually be taught. WYTIWYG might be paired in our thinking with the proverb, "Be careful what you wish for, because you may get it." The current emphasis on standardized testing of writing skills is likely to produce students proficient in the relatively simplistic skills they measure, and deficient in the more complex ways of thinking typical of Round Five in this study. If we desire an educational system that produces complex thinking, we will need complex assessments capable of discerning and rewarding it. I am hopeful that the concept of assessment communities will make some small contribution to the development of an alternative to the standardized writing test. It has the potential for conducting assessments that are far richer, more complex, more reflective of real writing work, and more valid than tests, and if further research supports these results, just as reliable. Being not only valid and reliable, but if brought to scale, less expensive than standardized tests, assessment communities could serve the purpose of accountability in a way that brings faculty to the center of the process rather than relegating them to the margins, and in doing so, produces faculty development that can be immediately utilized in the practice of teaching and assessing. This latter goal, improving the teaching and learning of writing, should be the first priority of all our assessment efforts.

Implications for Further Research

In the light of this discussion, I would first like to propose some ways further research could support or refute the usefulness of this type of Web-based activity for bridging the gap between high school and college in writing and other higher cognitive skills, especially with different populations of college and high school teachers, and then

to suggest some other educational problems for which assessment communities might be partial solutions.

Bridging the High School – College Gap

The findings of this project cannot be generalized to any other population, of course, but they raise several questions about how well this method of assessment would succeed in solving the vexing problem, noted in the introduction, of the misalignment between high school and college curricula. Many states today have started initiatives to align curricula between secondary and postsecondary schools with the aim of better preparing students for college-level work (Kirst & Venezia, 2004). As noted in the Literature Review chapter, however, standards and criteria for writing are very difficult to describe, and terms such as "clear," "consistent," and "effective," which are commonly used in curricula and scoring guides, fail to communicate with the degree of precision required for meaningful assessments – or teaching. The online assessment community has potential for operationalizing standards by using samples of student work, accompanied by commentary and analysis, to illustrate the various strengths and weaknesses expected at each score point. Although the problem of alignment is frequently conceptualized (usually by non-educators) as a simple matter of stating learning outcomes and lining them up in order, it is a far more difficult matter of communicating complex constructs that cannot be quantified and are not easily described at the level of specificity required for precise assessments. The most challenging learning outcomes – ones that prepare students for high-level work in the STEM fields (science, technology, engineering, and math), as well as in business, government, the military, the health fields, nonprofit work, academia, and the arts – cannot be described well, taught well, or assessed well on the

basis of printed sheets of paper. They require a live context of people, talk, social interaction, and reflection. Written down and shipped out of context, they lose not only their mooring, but their meaning.

The project described here shows some potential for bridging the high school/college gap in a meaningful way. It should be tested in many different contexts, with different groups, some similar to participants of this project and some quite different. Further studies are needed to determine whether similar effects could be obtained, especially with five populations different from ones studied here, and possibly others as well:

1) with less dedicated and knowledgeable teachers, both in the high school and college groups;

2) with different student demographics and levels of writing ability, both more advanced and less proficient;

3) with more demanding college curricula, such as those more focused on rhetoric or critical theory, or less demanding curricula such as those of developmental programs;

4) with larger groups of participants, where greater numbers of say, 25 or 100 participants online in the same discussion forum could weaken a sense of community, but provide access to more people.

5) with poor, urban schools that serve primarily minority students whose linguistic background and limited cultural opportunities create a gap between home and college far more formidable than the gap most students must negotiate. This could be the most difficult, and perhaps rewarding, trial of an online assessment community.

If other trials of the system are successful, strategies such as LINC may provide low-cost, accessible means of helping secondary schools align their writing curricula with college-level writing in ways that are more powerful than the publication of curriculum guides, with the result that high school students are better prepared for college work when they leave high school.

Trials with different designs. I would like to see how changes in the design of the study would affect outcomes. For example, if participating teachers were asked to respond to a checklist of items for each paper before joining the discussion forum, or required to post before viewing other posts, more evidence would be available for quantitative analysis of the discussion. Trials with fewer or more papers and with papers selected both for more variety and more uniformity of genre would probably produce different results, and trials with a different ordering of the procedures, such as viewing the PowerPoint slides before discussing the paper, would be interesting.

Results in teaching practice. What teachers learn in a summer workshop and what they are able to apply in their practice of teaching after school starts in September are two different things, of course. A study that follows teachers throughout at least one year of teaching after participating in this kind of assessment community would reveal whether new teacher knowledge is limited to posting in a discussion forum or is applicable in the classroom, and whether it is ephemeral or durable. Studies that examine student writing produced in experimental and control classrooms in the year after teachers participate in the study would be very useful in this regard.

Research into the "How" question. Learning how and why interpretive community forms among online groups of people is a good question for basic research.

Questions about the role of expressions of agreement and support; the amount, timing, context, and quality of expressions of disagreement; the amount and kind of discussion; and the role of reflective and introspective comments is just a beginning list of items that might be studied. Experiments with variations in the architecture of online interactions and planned injection of controversial issues into discussion might also further understanding. Answers to these fundamental questions about how community develops could be very important to the successful development of online communities in the future.

In addition to trials in bridging the high school/college gap with different populations, online writing assessment communities may have some usefulness in solving other problems in education. Some of the possibilities are discussed in the following sections.

Research in Other Applications

A large number of educational problems that are currently exasperating policymakers could be viewed as a lack of interpretative community among shareholders. Improving education in writing, critical thinking, and other higher cognitive skills depends in part on developing shared understandings and values about exactly what those constructs are, how they are recognized and responded to in practice, and what should be taught. Developing that understanding is not easily accomplished merely by publishing and disseminating documents, because the meanings of the typical terms of assessment (such as "pertinent," "rational," and "cogent") are, as Derrida (1976) would say, "slippery." They can easily mean different things to different people, especially with

regard to such questions as *at what level* or *to what degree* must a skill be demonstrated, and *how is that skill to be recognized* in the context of student work.

Traditionally, school districts have relied on workshops and inservice meetings to solve this problem, and most institutions of higher education have an office of faculty development that offers workshops and other supportive activities. Academic departments also sometimes engage in curriculum development and assessment initiatives that are effective in constructing the shared meanings that characterize interpretative communities, although their work is often viewed as a cognitive process based on *foundational* knowledge in a positive sense, rather than a social process of reacculturating learners into a new professional practice (Bruffee, 1999). Workshops and seminars are time-consuming and expensive, however, and as a result, they usually reach only small groups of people. Furthermore, they are often implemented as one-time events rather than continuing efforts, so that over time, especially with turnover in personnel, positive effects dwindle. Faculty development viewed as a social process of acculturating to a new community provides a more powerful metaphor for understanding how learning and change occur and designing appropriate activities to advance them. If further research confirms that online assessment communities are effective in building and maintaining interpretative community among large groups, they might be a cost-effective means of solving some of the following problems:

K-12 Faculty Development. Teachers can teach only what they know. The use of online assessment communities, with a sharp focus on the qualities and features of good writing and engagement in a substantial amount of analysis and scoring of student work, may present a low-cost, accessible alternative to workshops and lectures that usually

comprise K-12 professional development sessions. Just as in faculty development in higher education, learning among teachers at the elementary, middle, and high school levels can be understood as a social process of acculturation to a new professional community. Trials of online assessment communities at several grade levels, with scoring of pre- and post-intervention samples of student writing, would indicate whether a professional development effect occurs from participation in a LINC-type program in K-12 education. While not a substitute for the high-quality experiences such as the National Writing Project and similar intensive programs, the online assessment community may provide a substantial amount of faculty development within reasonable cost parameters. Additional research conducted at the elementary, middle, and high school levels could determine if teacher knowledge could be developed through online assessment communities at those levels.

Teacher education. It could be argued that assessment should be one of the most important components in a teacher education program, especially for those subjects such as English/Language Arts in which outcomes are so difficult to define and assess. Assessment focuses our attention on the qualities and traits of student work that we deem praiseworthy or blameworthy, and it raises to the surface our tacit assumptions and biases about what we really value (Broad, 2002). If teacher education majors learned to assess student work well, it is likely that they would be more astute in discerning positive and negative qualities of work and be capable of providing higher-quality feedback to students. The use of online assessment communities that connect experienced practitioners of assessment, such as classroom teachers with special expertise in writing, with undergraduate education majors could be a valuable, real-life component of

preparation for teaching. The opportunity to witness teachers discussing, analyzing, and evaluating student writing, to try their own hand at the practice, and to engage in collegial dialog with each other could provide a realistic teaching experience without even entering the classroom. The online experience, that is, could help induct new members of the profession into a community of practice, the practice in this case being assessment. A trial that linked teacher education students in English/Language Arts with practicing E/LA teachers in an online assessment community would test whether this application facilitates the induction of teachers-in-training into the professional community.

Standardizing K-12 learning outcomes in writing. Most states are attempting to standardize school curricula for all subjects and all grade levels within their borders to obtain more uniform results among various school populations. The two initiatives most frequently instituted for achieving standardization are publishing required curricula and instituting standardized testing. As noted earlier, descriptions of good writing and the typical scoring rubrics developed by state departments of education rely on terms such as "appropriate," "logical," and "evident" to communicate standards, and the terms alone are not helpful to teachers. They become meaningful when they are illustrated by examples of student work, accompanied by analysis and commentary that point out specific phrases and passages that exemplify the standard. Furthermore, although high-stakes testing has certainly increased teacher motivation, the test results are not always clearly understood in ways that lead to more effective teaching. The tests may actually have a deleterious effect on writing instruction, because they require practice in 20- or 30-minutes timed tests where writing is stripped of the context that makes it meaningful and complex (Huot 2003, Yancey & Weiser, 1999). Therefore, a program similar to

LINC should be tested to see whether teachers working in schools across a single state are able to develop clearer understandings of state standards, in lieu of standardized tests, that can serve as a guide to teaching and assessment practices that better advance students' writing skills. Again, when learning is viewed as a social process of forming community and inducting new members into the community, the online assessment community offers some opportunity for teachers to learn in more meaningful ways than through results of standardized tests. Pre- and post-measures of student writing would be required to determine whether participation in a LINC-type program aids uniformity and improvement of writing outcomes. It would be especially important to investigate whether an online assessment community could help bridge the gap in learning outcomes for writing between predominantly white middle class schools and poor urban schools where students of color are predominant. In this case pre- and post-measures would be compared across selected school districts.

Learning Beyond the Writing Curriculum. In a large sense, the two most-frequently mentioned higher cognitive skills, writing and critical thinking, are not separate activities but the same. Although writing includes some skills such as punctuation and basic grammar that are not considered critical thinking, and critical thinking includes some knowledge such as principles of formal logic that could not be considered writing, the two skills overlap far more than they diverge. It is difficult to think about complex subjects without committing thoughts to paper (or digital text), and writing well usually requires the use of logic, evidence, evaluation of sources, and so on. Furthermore, the set of skills we call critical thinking are typically both taught and assessed through writing. There are exceptions: moot court in law school, oral exams,

extemporaneous speeches, class presentations, and debate could all employ critical thinking without the use of writing. In most instances of the teaching of critical thinking, however, students are assigned papers to write, and the work is assessed in written form, making it assessable in an online community. Actually, any work that can be digitized could be assessed in an online assessment community, including photography portfolios; videotaped theatrical, musical, and dance performances; recorded speeches and debates; and so on. Any work that can be displayed online could be assessed and discussed by groups of faculty working collegially in online communities. In these kinds of learning experiences, as in writing, standards are difficult to communicate verbally in ways that are precise enough for assessment, but a small community of assessors who have developed shared values, language, and definitions could interpret and evaluate work reliably, in the convenient and accessible environment of the Internet. Trials using a program like LINC to assess critical thinking and other higher cognitive skills for disciplines other than writing could determine if online assessment communities would be useful for those purposes also.

Other applications. This study raises several questions about additional applications that can be answered only through further research. The BWC first-year writing courses focus heavily on the genre of the argumentative essay. Would teachers have experienced as much interpretive community if the genre had been narrative? It is not likely, but trials would be interesting. Another question is whether the teachers themselves would become better writers in their own right. It is generally acknowledged that many K-12 ELA teachers are not strong writers. Depending on when and where they were educated, even English majors may not have completed much challenging writing,

and many schools of education focus their elementary programs on reading. Furthermore, most K-12 teachers do not write professionally and therefore do not continue developing their writing skills after college. Could participation in online assessment communities improve teachers' own writing skills? This question would be easier to research than whether or not their students could write better a year later. Another critical question is how long the "norming effect" of participation would last. Typically, commercial and academic scorers "drift" away from established norms after the period of training ends. The question of how long the effect lasts has serious implications for estimating the long-term costs, of both monetary and faculty time, of maintaining a program. Finally, if the program has potential for both K-12 and undergraduate education, might it have some applicability in graduate education? Learning outcomes are extremely hard to specify and assess in the immensely diverse and complex realm of doctoral education. Some consideration could be given to a trial of online assessment communities among similar graduate programs. The Web platform would afford easy links across national or international distances.

Accreditation and Accountability in Higher Education. In the current political environment there is great pressure on higher education to demonstrate student learning in ways that are transparent and can be compared across institutions for the purpose of accountability. In an effort to deflect an attempt by pressures outside higher education to institute inappropriate measurements of outcomes, three of the six major associations of higher education, the National Association of State Universities and Land Grant Colleges (NASULGC), the American Association of State Colleges and Universities (AASCU), and the Association of American Colleges and Universities (AAC&U), have jointly

developed the Voluntary System of Accountability (2009) for institutions of higher education. The VSA has produced the College Portrait, an online portal that allows parents, students, and others to compare institutions on several measures. The most controversial measure is of learning outcomes. Currently, the VSA requires participating institutions to administer one of three standardized tests: the *Collegiate Learning Assessment (CLA),* the *Measure of Academic Proficiency and Progress (MAPP)*, or the *Collegiate Assessment of Academic Proficiency (CAAP).* As an alternative, however, the American Association of Colleges &Universities (AAC&U) is developing an online portfolio assessment called *VALUE: Valid Assessment of Learning in Undergraduate Education* (2009). When completed, it will offer colleges the opportunity to use assessments that faculty will likely find more valid and meaningful than the three available tests. LINC is similar to the AAC&U portfolio initiative, except that it places greater emphasis on collegial participation, practice, and dialog, and it does not necessarily require a rubric. Also, I envision a process of developing a national assessment system that begins with hundreds of local assessment communities that establish high local validity, and later institutes a gradual merging of local assessment communities with peer institutions, working toward a national standard. The AAC&U effort started with a fairly small group of people and has been gradually expanding it to larger and larger groups. The AAC&U VALUE project and the LINC project studied here move in opposite directions (central authority to local; local authority to central) but nevertheless work toward the same goal, eventually developing national assessments of higher level skills that are both valid for the purposes of improvement and at the same time reliable enough to meet the calls for accountability of student learning outcomes.

The key to validity in these efforts, I believe, is the development of shared values, perspectives, and language that results from the kind of formation of interpretative community demonstrated in this study. It is important to recall, in this respect, Bruffee's (1999) explanation of how collaborative learning results not from abandoning previously-held values and ideas, but by nesting them within the larger perspectives of the new community. As each local group merges its assessment activities with another group, another round of sharing, discussing, analyzing, questioning, challenging, modifying, and learning begins. Each group influences the other, not by replacing old ideas with new ones, but by collaboratively creating new knowledge – a synthesis of old ideas that is a higher, more complex level of understanding than the old ones, but still connected to them.

Without the development of complex, nuanced, organic social knowledge that is generated, maintained, and continually modified in professional communities, neither the results of standardized tests nor the distribution of curricula and rubrics is likely to be effective at bringing about real change in education. Trials that establish local online assessment communities at different institutions and then gradually link them with each other to form a larger community would be a large undertaking, but the results could possibly point to a means of assessing higher cognitive learning outcomes with a high degree of validity, while also affording the opportunity to compare results among similar institutions. Reports to the public would consist of percentages of students scoring at each score point on a common scale, with the standards continually refreshed and updated by the work of local communities. The same kind of effort may help states produce

meaningful, high-level assessments of writing and critical thinking for accountability purposes in K-12 education.

Implications for Practice and Policy

This small study raises some questions about how assessment of student learning outcomes can best be achieved. It should be noted that the learning outcomes under consideration in this entire study are higher-level cognitive skills, especially in the area where writing skills and critical thinking skills overlap. Many other learning outcomes are related to the specialized content knowledge of the disciplines, such as anthropology and astrophysics, or to the kind of skills where a single right answer or best answer is evident, as in the STEM fields. Clear communication about exact requirements of student learning is much easier to attain in these areas, as is meaningful assessment of student learning. Those areas are not a concern here. It is higher-level skills that require divergent, original thinking, such as writing and critical thinking, that require the more advanced means of communication and assessment afforded in online assessment communities. Separate sections of implications follow for administrators and for policymakers.

Implications for administrators of postsecondary institutions

In light of the findings of this study, it would seem wise for administrators to hold out against political pressures for standardized testing at the postsecondary level for the time being, and to encourage trials of local assessment communities where possible. This would be especially true for administrators of first-year composition courses and those of other general-education programs whose goal is to introduce students to the intellectual

rigor of higher learning and to develop skills in critical thinking and communication. The learning outcomes of these programs are best understood not as knowledge existing independently as an external reality that students must acquire, but as socially-constructed knowledge that is best learned in a community of learners. Utilizing the socially-mediated contexts of writing and thinking to assess learning outcomes, rather than stripping away all context as required by standardized tests, is an appropriate – and possibly powerful – way to assess learning outcomes with very high validity and at least acceptable levels of reliability.

Furthermore, local online assessment communities are not difficult to develop or expensive to maintain: the technology is relatively simple, and faculty already have the knowledge and skills needed to build the content. Although some costs are entailed in the initial development and in maintenance, the costs would be quite low on a per/person basis when the project is extended to large groups of people. Faculty learning that is immediately and directly applicable in teaching practice is far more likely with this type of assessment than with standardized tests. Considering that local assessment communities have great potential for solving not one, but two problems (faculty development and assessment), the expense and administrative effort to establish them seems quite reasonable, especially when compared with the costs of administering both standardized assessments and faculty development programs.

Implications for policymakers

The results of this study should be a caution signal to policymakers who have strong faith in standardized tests for accurate assessment of learning in higher cognitive skills. A fairly quick examination of the depth and complexity of the analysis of student

writing performed both by the college teachers throughout this study and by the high school teachers in Round Five, should be compared with the level and type of thinking required for standardized writing tests. Particularly, the depth and detail of the analysis of papers in this study should be compared with the scoring regimens used by commercial tests, including the analysis of sample papers and rubrics used to train scorers. It is impossible for standardized tests to achieve the scope and depth of the writing produced by the range of first-year students who volunteered their papers for this study, because tests preclude the skills of locating, evaluating, and incorporating source material into original and complex theses that synthesize diverse ideas. Even the rudimentary paper that received the lowest score of 0.0 in this study, Paper 3.5 on why celebrities make poor role models, synthesized several sources into an original thesis, utilized source material as evidence, and drew conclusions from evidence. That level of thinking is superior to the demands of standardized writing tests, which require students only to support whatever opinions they already hold with whatever evidence comes to mind during the test period.

The result of a comparison between analysis of papers written for standardized tests and analysis of papers conducted for this study will almost certainly show that standardized writing tests do not require, and do not measure, nearly the level of detail, complexity, logic, rhetoric, and other qualities that the first-year composition teachers in this study require in their courses. A 20-minute test, or even one of twice that length, can only skim the surface of the type of writing and thinking skills expected of students by the college teachers in this study.

I would like to point out that, although I consider the "Blue Water College" selected for this study to be an exemplary community college, it is by no means selective.

It is an open-admissions institution that serves a population of largely rural and small-town residents in the northern reaches of a Midwest state. If the requirements of complexity, logic, development, use of source material, and so on are steep at this school compared with the requirements of standardized tests, it is likely that the requirements of many other colleges and universities across the country would be even more so.

A relatively small investment in research on online assessment communities could potentially lead to the development of a means of assessment, development of national standards, and communication with the public about learning outcomes that have far more construct validity and just as high reliability as standardized tests, and at the same time produce valuable faculty learning, at a relatively low cost. Three types of funders could potentially be interested in the project. First, several of the major philanthropic foundations, such as Lumina, The Bill and Melinda Gates Foundation, the Teagle Foundation, and the Spencer Foundation, are currently supporting efforts at school reform, all with a heavy emphasis on assessment and accountability. Second, the Obama administration has posted this language on the White House education policy page:

> Obama and Biden believe teachers should not be forced to spend the academic year preparing students to fill in bubbles on standardized tests. They will *improve the assessments used to track student progress to measure readiness for college* [italics added] and the workplace and improve student learning in a timely, individualized manner.

This project could potentially provide improved assessments for college readiness in higher cognitive skills that are more meaningful than current measures, and the federal government currently has both the motivation and the resources to investigate new methods. Third, it is possible that departments of education in some of the 50 states could also be interested in funding smaller projects within their boundaries that test the

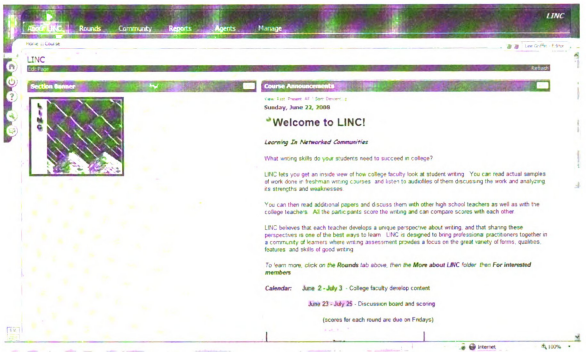
effectiveness of online assessment communities in raising college preparation in the higher level skills.

In the spring of 2009 as this research is concluded, the Voluntary System of Accountability is a few months away from publishing the results of a study on the construct validity of the *Collegiate Learning Assessment (CLA), Measure of Academic Proficiency and Progress (MAPP), and Collegiate Assessment of Academic Proficiency (CAAP)* tests; and the e-portfolio assessment project of the AAC&U is coming to fruition. A new U.S. Secretary of Education has been invested with the responsibility of spending over five billion dollars for reform initiatives, and a new administration in Washington has pledged to make education one of the main pillars of its policy to transform the economy. There could be no more advantageous time for harnessing a new technology to serve student learning – and no more dangerous time to ignore new technologies in favor of the conventional practices of the commercial testing industry. Policymakers have an opportunity in this rare moment of national crisis to attempt some new solutions as suggested by the results of this study. Much is at stake.

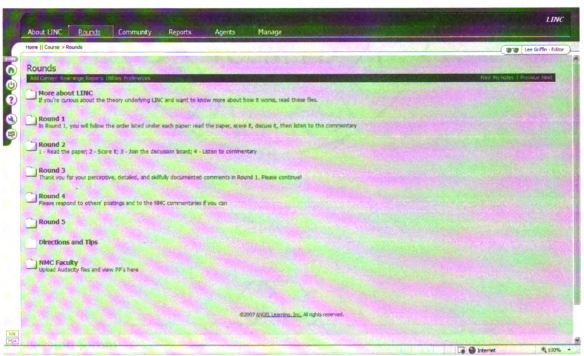
APPENDICES

Appendix A: Screen Shots of LINC Web site

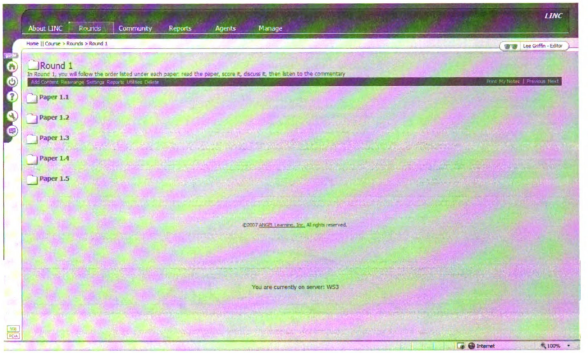
Welcome page:



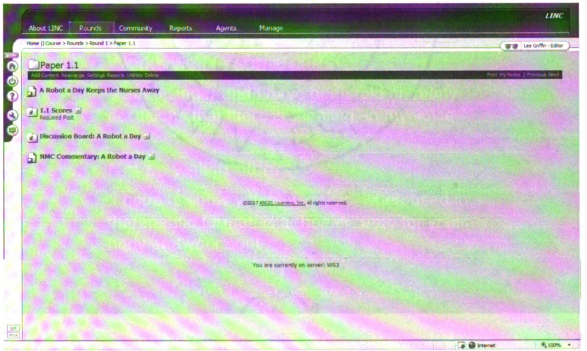
The Rounds:



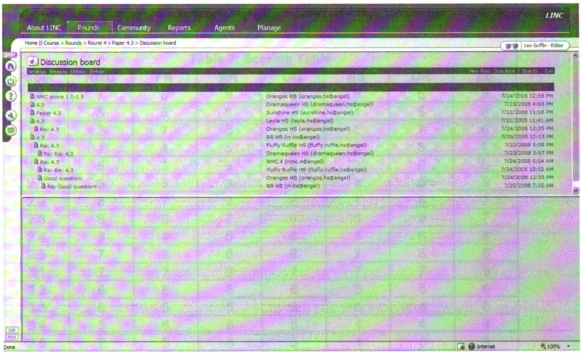
Papers in Round 1:



Activities in Paper 1.1:



Discussion forum:



Slide in PowerPoint for Paper 1.3:

Appendix B: Scoring Table

Table 7: Scoring Table

Paper	NMC	FR	RR	DQ	SS	OR	LL
1.1	7	6	5	6	5	5	8
1.2	5	5	8	7	4	6	6
1.3	7	7	8	7	5	7	6
1.4	1	4	5	6	6	6	7
1.5	5	3	6	5	3	6	0
2.1	9	6	9	7	8	6	9
2.2	8	7	9	8	7	8	7
2.3	8	6	7	6	6	9	7
2.4	3	4	5	6	4	6	6
2.5	7	8	6	8	7	8	8
3.1	7	7	9	7	7	7	9
3.2	6	6	6	6	5	5	6
3.3	6	5	5	6	6	4	6
3.4	5	6	6	5	4	4	4
3.5	2	5	5	5	4	5	5
4.1	8	7	9	8	8	8	9
4.2	3	5	5	6	4	2	3
4.3	4	6	3	4	5	2	6
4.4	8	8	8	8	8	6	9
4.5	4	5	5	6	3	7	5
5.1	5	5	5	6	3	7	5
5.2	8	9	9	9	9	8	9
5.3	9	7	9	8	9	9	9
5.4	7	4	6	5	6	6	3
5.5	4	3	5	5	2	7	5

Appendix C:
Initial High School Interview Protocol

Thank participant for agreeing to participate.
Summarize Human Subject rights and emphasize right to decline or withdraw or refuse to answer any question.
Remind participants that data may be shared with Principal Investigator and Dissertation Committee.
Summarize purpose and procedures of study.
Double-check permission to record interview on audiotape for purpose of accuracy.
Ask if participant has any questions before beginning.
(Note: Only Group C and Group D questions will be used in the analysis. The others are used to set the participant at ease and gradually move the conversation to specific issues, and they may have explanatory or contextual value.)
Group A: Questions to verify demographics:
1. How long have you taught at _____ High School?
2. Have you taught expository writing/college-prep writing a long time?
3. What other courses have you taught?
4. What degree(s) do you hold, and from where?
5. What was most helpful to you in learning to teach writing: *(Ask as open-ended question first, then prompt)*

- Teacher education courses
- English or related courses
- Curriculum guides in your school
- Discussions with other teachers
- Professional conferences and publications
- Support from a department head or director of instruction
- Textbooks used in your classroom
- Your own experience
- Other?

6. Roughly, what percentage of your 12th-grade students goes on to college, and where do they go?

Group B: Questions to ascertain the teacher's present practice in teaching writing and develop a framework for more specific questions to follow:
7. What traits or qualities of writing are taught in your curriculum? (Give examples to prompt if necessary: personal essay, book report, research paper)
8. What genres or types of writing do you assign?
9. About how much writing do your students over the course of a semester (verify that a semester is roughly 18 weeks and half the school year)?
10. What are the biggest problems you see in the writing of 12-grade students, in terms of their preparation for writing in college?
11. What lessons or strategies do you find most helpful in improving student writing?

12. What resources do you use frequently to teach writing (textbooks, manuals, etc.)?

Group C: Questions to identify the teacher's initial understanding of the expectations of college writing instructors:
13. What qualities or traits of writing do you think will best prepare your students for the writing they will do in college? *(Ask as open-ended question first, then prompt from the following list. The list may be modified or expanded after the work of the college instructors and again during the course of the interviews.)*

- Conventions of spelling, punctuation, capitalization, format
- Organization and structure, including "sticking to the point"
- Originality, invention, creativity
- Authentic personal voice
- Grammatical correctness
- Expressiveness; ability to evoke images, moods, or feelings
- Ability to frame and develop a logical argument
- Awareness of audience and ability to address it
- Knowledge of specific rhetorical strategies
- Ability to write quickly, within specific timeframes
- Willingness to revise extensively
- Other?

Group D: Questions to identify the relative value the teacher assigns to various traits:
14. What traits or qualities do you think are most important to develop in writing of 12-grade students? *(Ask as open-ended question first, then prompt from the following list. The list may be modified or expanded after the work of the college instructors and again during the course of the interviews. The entire list probably will not be repeated depending on the context and flow of discussion, but is included here as a prompt for the interviewer.)*

- Conventions of spelling, punctuation, capitalization, format
- Organization and structure, including "sticking to the point"
- Originality, invention, creativity
- Authentic personal voice
- Grammatical correctness
- Expressiveness; ability to evoke images, moods, or feelings
- Ability to frame and develop a logical argument
- Awareness of audience and ability to address it
- Knowledge of specific rhetorical strategies
- Ability to write quickly, within specific timeframes
- Willingness to revise extensively
- Other?

Group E: Questions to identify the teacher's prior experience for and attitude about participating in the study:
15. Have you ever gotten information from any college instructors, about the expectations they have incoming first-year students? If so, please describe. (Prompt if necessary: talked to them, had meetings with, heard at conferences, read material.)
16. Have you ever participated in an online learning experience? If so, please tell about it, including the quality of your experience and how worthwhile you thought it was.
17. How confident do you feel about using the Internet as a learning tool and about navigating around Web sites with posted documents and discussion boards?
18. Do you have a reliable, high-speed Internet connection at school? At work?
19. Is there anything I didn't ask about that you think would be relevant to this study?
20. Do you have any questions before your first session?

Verify time and place of the first scheduled session.
Remind teacher that discussion boards may include strong feelings or controversial issues about student writing; that discussion is expected to remain professional and collegial; and that they have the right to withdraw at any point without consequence.
Make sure they have contact information in case of problems.
Thank the teacher again for participating.

Appendix D:
Mid-Project High School Interview Protocol

Thank teacher for participating.
Congratulate him/her on reaching the midpoint of the project.
Ask if the general tone and nature of the discussions has been appropriate and comfortable; if not, remind participant they can withdraw at any point without consequence.
Remind participants that data may be shared with Principal Investigator and Dissertation Committee.
(Note: The initial questions are not designed primarily for analysis, although they may provide valuable context. They are intended to start the interview on a positive note, put the teacher at ease, address outstanding questions quickly, and introduce the main issues gradually.)
Group A: Question to start the session positively:
1. Tell me generally – how is it going?
2. Are you having any technical problems? If so, please describe.
3. Are you having any trouble understanding the college instructors' comments or scoring methods?
4. Are there any other aspects of the project you find troublesome or difficult?

(Address any outstanding issues as directly and promptly as possible; promise to investigate and support if not able to solve problem immediately.)
Group B: Questions to explore teacher's reactions to college instructors' analysis and rating of student writing:
5. Have you found so far that there are too many student papers to consider, not enough, or about the right number of papers?
6. Thinking back to the sample papers you have read and the comments of college instructors about them, how often did you find their comments and ratings in accord with your own thinking, and how often were you surprised by what they said? Please give examples when you can.
7. On what traits or qualities of student writing did you find yourself in close agreement with college instructors?
8. On what traits or qualities of student writing did you find yourself in low agreement with college instructors?
9. Have the college instructors made comments that you found confusing, ambiguous, or even erroneous? *(Prompt with examples from the discussion board if items have appeared there.)*
10. Judging from their comments and ratings, what do think college instructors value most in student writing? Did you find this similar to your own values of writing, or are their values quite different from yours?
11. *(If there is a high differential between college instructors' valuing of writing traits and the teacher's own)* Is this difference likely to be problematic for you in your teaching, or do you think you will easily take their perspectives into account and incorporate them in your teaching? Please explain.

12. Have you been surprised at anything you've read or heard from the college instructors on this Web site?
13. How closely do the comments of other 12th-grade teachers mirror your own thoughts about the writing samples? On what points do you most agree? On what points do you most disagree?
14. Have you noticed any ways in which your thinking about traits and qualities of student writing has changed since you started the project? If so, please describe.

Verify timeline for completing the project.
Verify time and place of next interview.
Make sure they have contact information in case of problems.
Thank the teacher again for participating.

Appendix E:
Final High School Interview Protocol

Thank teacher for completing project.
Remind participants that data may be shared with Principal Investigator and Dissertation Committee.
Congratulate him/her on reaching the end of the project.
(Note: The initial questions are not designed primarily for analysis, although they may provide valuable context. They are intended to frame the interview, put the teacher at ease, and introduce the main issues gradually.)
Group A: Questions to begin the interview.
1. Give me a general overview of your thoughts on your experience in this entire project.
2. What have you most enjoyed about the project?
3. What, if anything, has caused stress, annoyance, discomfort, or any other negative reactions?
4. Have you found that there were too many student papers to consider, not enough, or about the right number of papers?

Group B: Questions to ascertain learning from the project.
5. Thinking back to the interview we did just before you started the project, can you identify any specific ways your thinking about student writing has changed? *(If necessary, prompt with a few quotations from the initial interview.)*
6. Are there any traits or qualities of student writing that college instructors place a higher value on than the usual high school curriculum does?
7. Are there any traits or qualities of student writing that college instructors place a higher value on than your own teaching does?
8. On what items did you agree with the college instructors?
9. Thinking about all the comments you exchanged with other teachers of 12^{th}-graders on the discussion board, were your views very different from most of the them at the beginning? At the midpoint? By the end of the project?
10. Please describe any movement you observed in your own comments on the discussion boards, and in the comments of others.
11. Did you observe any changes in the way you scored student writing over the course of the project?
12. Did you observe any changes in the way other teachers scored student writing over the course of the project?
13. Are the scores of student writing you assigned at the end of the project closer to that of the college instructors, about the same in how much they differed, or even further apart?

14. *(If the teacher observed differences)* To what would you attribute those differences?
15. Are the scores of student writing you assigned at the end of the project closer to that of the other 12^{th}-grade teachers, about the same in how much they differed, or even further apart?

16. *(If the teacher observed differences)* To what would you attribute those differences?

17. Is it likely that your participation in this study will affect your teaching of writing to 12^{th}-grade students in the future? If so, please describe how.

18. Considering the amount of time you invested in this project and your overall experience with it, do you think it has been worthwhile for you? Please explain.

19. If this Web site is made available to other teachers of high school English, would you recommend they participate? Why or why not?

20. Are there technical, navigational, or graphical features of the Web site you think should still be modified before offering it to additional teachers of writing?

21. Are there aspects of the study I haven't asked about that you feel are relevant to report? Please explain.

22. If I carry this work forward, would you make any recommendations for improving the Web site, the methods used in the study, or anything else?

Affirm that I will verify the accuracy of his/her comments as they are transcribed.

Give an expected date of completion of the study and promise to send an electronic copy if the participant desires.

Thank the teacher again for all the time and effort devoted to the study.

Appendix F:
Final College Interview Protocol

Thank instructor for participating and seeing the project through to completion.
Remind participants that data may be shared with Principal Investigator and Dissertation Committee.
(Note: the first question is designed primarily to begin the interview positively, refresh the participant's mind about previous sessions, and introduce the main issues gradually, but they may have explanatory and contextual value in the analysis.)

1. You began this project last March *(or whenever).* Since that time, your group has produced a lot of material. Has it been generally a positive, negative, or mixed experience?
2. What was the most difficult part of the project for you? *(Ask as an open-ended question first, then prompt if necessary. The list may be modified or extended for subsequent interviews.)*

- Deciding on sample papers
- Agreeing on what features of the writing were salient
- Understanding diverse perspectives and value schemes
- Agreeing on common scores
- Other?

3. In the course of developing material for the Web site, you had several meetings with other instructors in your department and shared your perspectives and insights about student writing. What were the positive aspects of that experience? The negative?
4. Do you think the collaboration required for this project will influence your assessment of student writing in the future? The assessment of writing by the other participants? Assessment in your department?
5. You've had a chance to see how your work is displayed and utilized in the Web site. What are your general impressions of the way teachers responded to your work?
6. What are your general impressions of the way teachers interacted with each other on the discussion boards?
7. You've also had a chance to see how 12^{th}-grade teachers of writing responded to the sample papers. What were your impressions of their discussions? Did anything surprise you, or could you have anticipated most of their commentary?
8. How useful, overall, would you say this kind of communication via an interactive Web site is for developing shared understandings of writing assessment between high school and college instructors?
9. If I carry this work forward, what recommendations would you suggest for improving the Web site, the method of study, or any other aspect of the project?
10. Is there anything I didn't ask about that would be helpful in my analysis of the project?
11. On balance, considering all the time and energy you and others put into this study, do you feel it will have a positive effect on the ability of participating high school teachers to prepare their students for college writing?

12. Knowing what you know now, if you were asked to participate in a similar study in the future, would you be inclined to accept or decline?

Give an expected completion date for the study and promise to email a copy if he/she desires.
Invite him or her to stay in touch for future developments.
Thank instructor again for the time and effort devoted to the study.

Appendix G: Table of Scores for Individual Interrater Reliability Correlations

Table 8: Table of Scores for Individual Interrater Reliability Correlations

Round	FR	RR	DQ	SS*	OR	LL
1	0.781	0.490	0.250	-1.110	0.000	0.000
2	0.738	0.909	0.526	0.903	0.556	0.769
3	0.656	0.722	0.743	0.867	0.500	0.763
4	0.853	0.937	0.835	0.972	0.841	0.945
5	0.835	0.963	0.875	0.930	0.674	0.800

*The text box below is an explanation for negative correlations by Garson (2009). The negative datum for SS in Round 1 is mostly likely explained by small sample size, described in the boldface type below, since there is no problem with data entry, measurement, or multidimensionality.

How should a negative reliability coefficient be interpreted?

It should be interpreted as a data entry error, a data measurement problem, as a problem based on small sample size, or as indicative of multidimensionality. As negative reliability is rare, the researcher should first check to see if there are coding or data entry errors.

One situation in which negative reliability might occur is when the scale items represent more than one dimension of meaning, and these dimensions are negatively correlated, and one split half test is more representative of one dimension while the other split half is more representative of another dimension. As Krus & Helmstadter point out, factor analyzing the entire set of items first would reveal if the set of items is plausibly conceptualized as unidimensional.

A second scenario for negative reliability is discussed by Magnusson (1966: 67), who notes that when true reliability approaches zero and sample size is small, random disturbance in the data may yield a small negative reliability coefficient.

In the case of Cronbach's alpha, Nichols (1999) notes that values less than 0 or greater than 1.0 may occur, especially when the number of cases and/or items is small. Negative alpha indicates negative average covariance among items, and when sample size is small, misleading samples and/or measurement error may generate a negative rather than positive average covariance. The more the items measure different rather than the same dimension, the greater the possibility of negative average covariance among items and hence negative alpha.

Appendix H: Definitions of Dimensions

Classifying the comments on the discussion board was difficult because there is no widely-accepted taxonomy for the terminology of composition instruction and assessment. Many terms in common use have several meanings, and categories typically overlap. For example, when a student writes "there" instead of "their," is that a spelling error, a homonym error, a usage error, or a grammatical error? Different professionals describe the error in different ways.

The description of higher-level cognitive skills is similarly confusing. For example, a young writer may switch abruptly from one paragraph to another without making the relationship between two ideas clear. That kind of error can be perceived and described as a problem of organization (arrangement), as a lack of coherence, as a need for transition language, or as faulty logic. Different teachers may perceive the problem differently, and therefore prescribe different remedies.

For the purposes of this study, it was necessary to make some decisions about how writing problems should be categorized so that comparisons could made from one participant to another, and from the beginning of the study to the end. The following table presents the resulting categorization. It may not be useful to composition and rhetoric instructors generally, but the categories adequately capture the comments of the participants and they afford a consistent and coherent classification of the diverse terminology used in the transcripts.

Not incidentally, the exercise of developing and applying these categories underscored the importance of Stanley Fish's assertion that the way we see things determines how they are "made." It also echoes the difficulty of bringing order and

standardization to the splendidly diverse realm of thought among faculty who teach higher-level cognitive skills.

Table 9: Definitions of Dimensions of Writing		
Feature	**Dimension**	**Definition**
Topic	Relevant/Resonant	The topic is relevant and current with times; it resonates with a general adult audience; it would be considered important or interesting by most adults.
	Suitably narrowed	The topic is large and general enough that the writer has plenty to write about, but narrowed enough to allow the writer to address it in detail rather than superficially.
Introduction	Engaging	The introduction includes language or rhetorical devices or strategies to pique the reader's interest rather than using prosaic language and a simplistic statement of what the paper will cover.
	Apt	The introduction serves to direct the reader's attention to the topic of the paper, provides adequate background information, and helps the reader "enter the writer's world."
Thesis	Clear	The reader can discern quickly what the thesis is, or if it is tacit or delayed, the reader is not confused during the reading about where the paper is going.
	Complex	The thesis does not address the issue in a simplistic black-and-white fashion, but acknowledges opposing viewpoints or contradictory information, and is appropriately qualified.
	Focused	The thesis is not overly general and vague. Precise, specific language is used to convey not only what the paper will cover, but what it will not cover.
Argument	On point	The flow of argument from beginning to end of the paper is directed to the main point conveyed in the thesis, and does "ramble" into irrelevant subtopics.
	Logical/Coherent	The argument from beginning to end flows in an easily-followed line of reasoning. Conclusions and interpretations are consistent with evidence presented and principles of reasoning. The writing avoids leaps of reasoning, non sequiturs, and unsupported generalizations.

Table 9 continued

	Explicit	The relationships between evidence and conclusions drawn from it; between various parts of the argument and each other; and between parts of the argument and the thesis must be clear. The writer does not leave writers to make inferences or connect relationships on their own.
	Counter/concession	The writer acknowledges weaknesses in his/her argument, incomplete or contradictory evidence, valid points made by opponents, and multiple viewpoints. Points are conceded and conclusions are qualified when appropriate.
	Depth/complexity	The writer does not stop at superficial, ideological, or cliched knowledge, but probes to gather additional information and presents an understanding of multiple perspectives.
Support	On point	The evidence presented from print and digital sources directly supports the thesis and is related to the context in which it is used.
	Fully developed	Source material is not just mentioned or quoted, but is explained in detail and interpreted when appropriate. Arguments proceeding from logic are fully explained.
Sources	Support claim	Source material is directly related to the thesis and bolsters the argument. Evidence is not "thrown in" just to look good.
	Quality	Sources are primarily from quality sources, such as peer-reviewed journals, academics associated with prestigious institutions, and media which maintain high journalistic standards. Internet, radio talk show, and popular media sources are evaluated carefully or eschewed altogether.
	Variety	The writer synthesizes material from a variety of sources rather than overusing one primary source, which generates a "book report" type of writing.
	Integrated smoothly	The sentences leading up to and following material that is directly or indirectly quoted are complete, balanced, and rhythmic; quotations are not just "dropped in."
Citations	Correct	In-text citations and the Works Cited page (or Works Consulted page) are formatted in MLA style.

Table 9 continued

	1-to-1 with references	Every source cited in the essay is included in the Works Cited page, and every source in the Works Cited page is cited in the text.
Voice	Consistent	The writer maintains a clear stance and persona throughout the essay, rather than switching, for example, between academic and personal voices or adult and teen voices. Also, the writer maintains 3rd person pronouns in formal writing and avoids the second-person "you" and reference to first-person "I" and "me" in formal writing.
	Authentic	The writer's stance toward his/her subject and conviction for the argument appear to be genuine and committed.
	Appropriate	The writer maintains more formal, academic language in formal writing and utilizes a more informal, personal voice in personal narratives and essays.
Rhetoric	Devices/Strategies	The writer employs a variety of rhetorical devices (such as the rhetorical question or a metaphor eliciting a common experience) or rhetorical strategies such as the classical pathos, logos, and ethos, to make an argument effective with a particular audience.
Rhetoric, cont.	Arrangemt/Transitions	Often referred to as "organization," arrangement is the order and placement of various parts of the argument in the essay. It is difficult to separate from "transitions," because there may be many ways to arrange subpoints or episodes if they are appropriately related to each other. Transitions include both direct ("Now I will explain...") and indirect methods. Indirect transitions can be effective if the reader follows the argument without confusion.

Table 9 continued

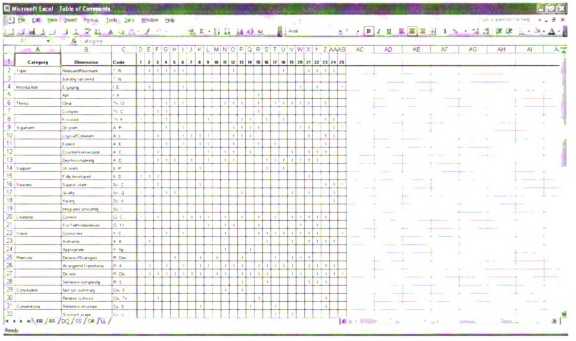
	Diction	Specifically, diction refers to word choice. Over the course of an essay, word choice is instrumental is developing a particular style and voice. Academic terminology adds to the formality and authority of piece, for example. Also, the writer avoids slang and cliched terms for scholarly writing.
	Sentence complexity	Grammatical complexity itself is not necessarily a goal; rather, grammatical complexity appropriately employed to convey complex meanings and relationships are valued. A variety of sentence constructions used with balance and rhythm is valued.
Conclusion	Not just summary	The conclusion does not simply summarize the argument. It may extend the meaning, raise questions, look to the future, or suggest additional applications, for example. A concluding thought that is more philosophical or inspirational is appreciated, as long as it is consistent with the general thrust of the paper.
	Relates to thesis	The conclusion, however original or literary, is expected to bolster the thesis in some way, although it may do so indirectly.
Conventions	Sentence structure	Both simple and complex sentences must be correctly formed and ended with a period. Fragments, comma splices, and run-on sentences are counted here rather than in the punctuation category.
	Standard usage	The writer avoids nonstandard constructions such as "We was..." and "She don't..." The use of slang and failure to use gender-neutral language is counted here.
	Spelling & homonyms	This category includes standard spelling, and also homonyms such as "there," their," and "they're", and close homophones such as "accept" and "except."
	Punctuation & capit.	All punctuation other than end punctuation for sentences is included here. Also included is the proper use of capital letters. Errors in punctuating quoted source material are included here.

Table 9 continued

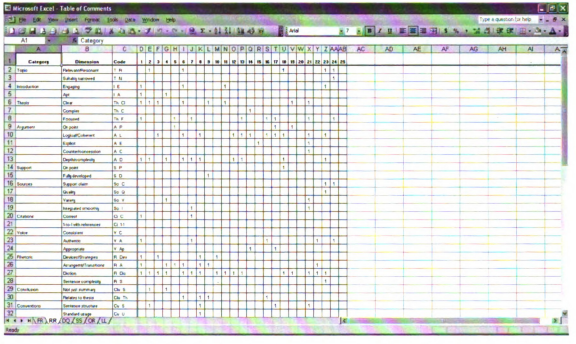
	Grammar	This category includes subject-verb and number agreement, shifts in verb tense, dangling modifiers, and similar errors not included in previous categories.
	General	Occasionally, participants commented generally on "errors," "lack of proofreading," etc. Those comments are tallied here.

Appendix I: Tables of Comments

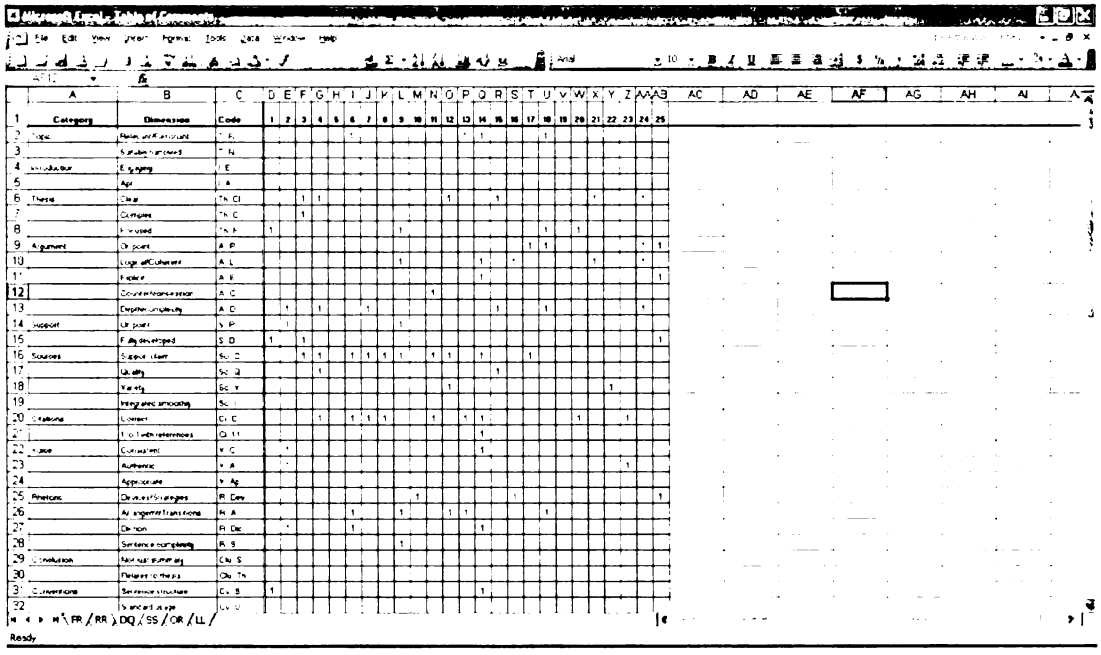
FR's Table of Comments:



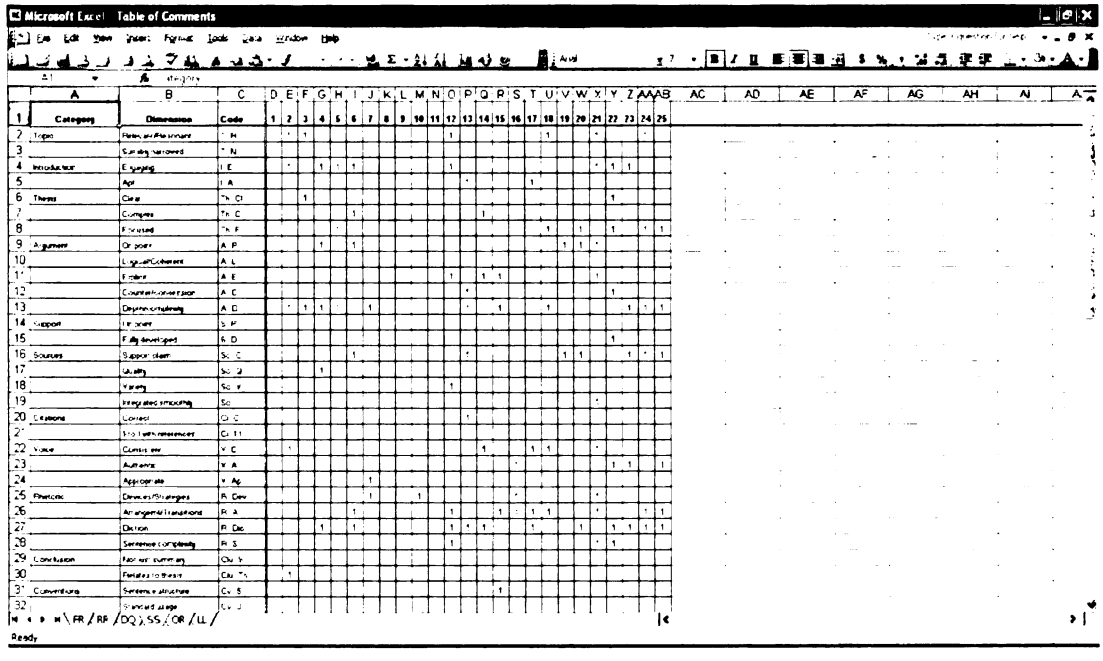
RR's Table of Comments:



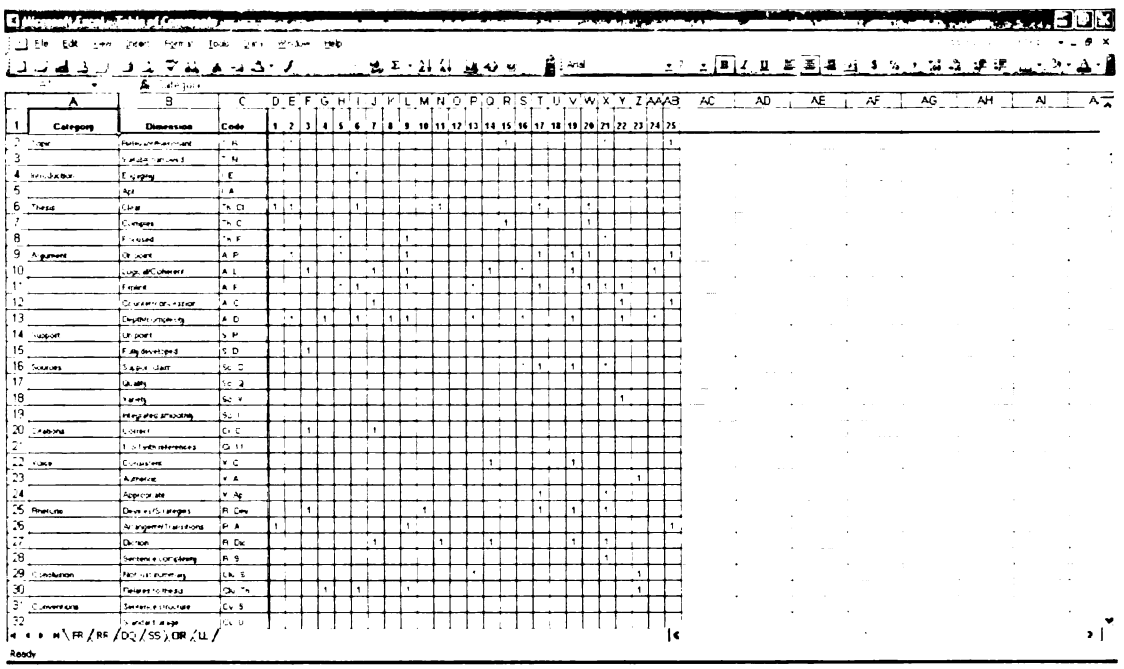
DQ's Table of Comments:



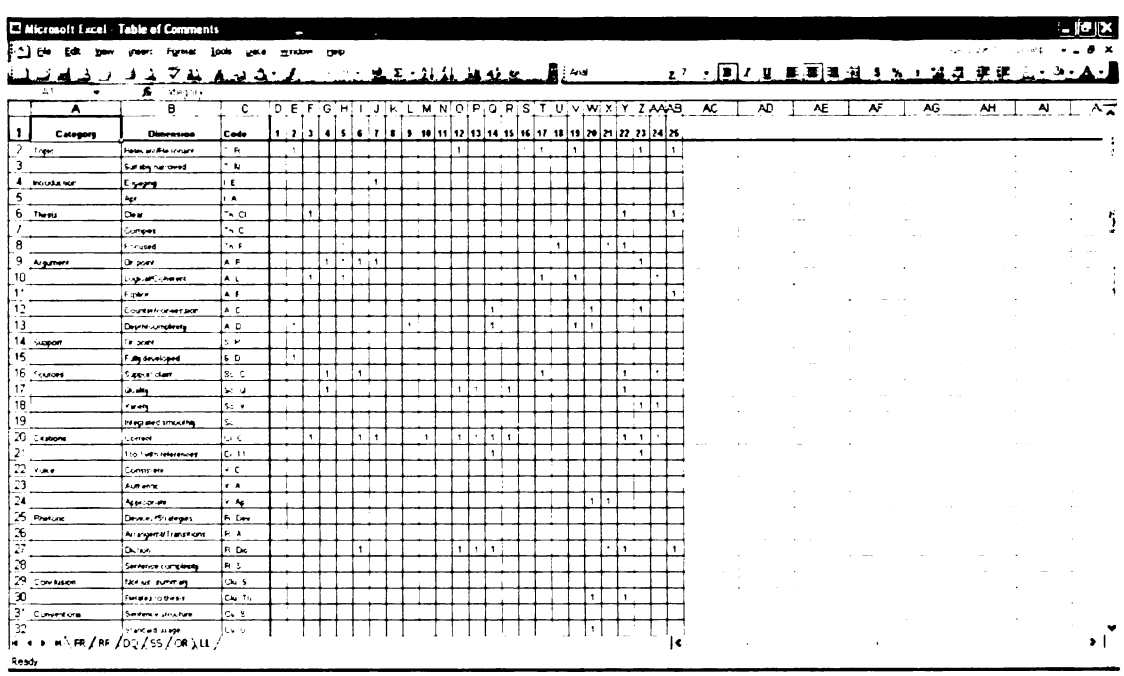
SS's Table of Comments:



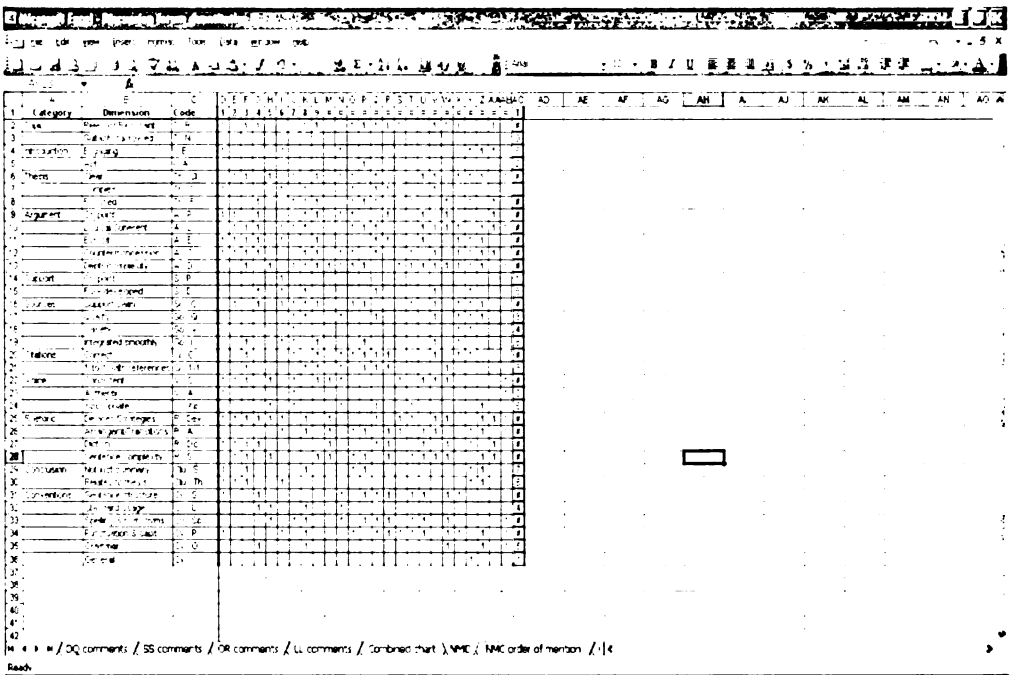
OR's Table of Comments:



LL's Table of Comments:



The college teachers' Table of Comments:



Appendix J: Notes on Tabulation of Comments

These notes are intended to explain some of the decisions made about the numerical tabulation of discussion forum comments that were coded for analysis.

Originally, each Dimension was coded either a "1" to indicate its presence in the writing (as in, "This paper contains a counterargument") or to indicate a positive evaluation of the item (as in "The counterargument is strong"). A negative one ("-1") was used to indicated the absence of a feature ("The paper lacks any counterargument") or a negative evaluation ("The counterargument is weak"). Eventually, this system proved ineffectual. Many of the teachers' evaluations fell somewhere near the midpoint of the scale, neither very positive nor very negative, so that two evaluations that were quite similar ("The counterargument needs just one more example to be effective" and "Although the counterargument would improve with one more example, it is still effective") would result in a completely opposite score in the tabulation. Another difficulty with this method is that frequently, writers do something well in one part of a paper and poorly in another, so that commenters would make both positive and negative statements about a single dimension within a single paper.

This problem was solved by returning to the way the research question was framed by Fish (1980): "...[M]embers of the same community will necessarily agree because they will see (and by seeing, make) everything in relation to that community's assumed purposes and goals: and conversely, members of different communities will disagree because from each of their respective positions the other "simply" cannot see what is obviously and inescapably there (p. 15)."

The focus, then, turned to whether or not the participant "saw" the feature in the writing and felt it was worth commenting on, rather on how positively or negatively it was valued. This proved to be an effective way to capture the participants'changing perceptions and valuing of features in student writing, especially when compared with comments made previously by the college faculty. Therefore, the final table displays a "1" to indicate whether or not the participant commented about that dimension for that paper.

Appendix K: Definitions of Perspectives on Writing

Table 10: Definitions of Perspectives on Writing

Perspective	Definition
Pedagogical Approach	The teacher framed criticisms of the paper not in terms of good/bad or right/wrong, but in language that would be helpful to the writer in understanding his/her writing and in revising it.
Awareness of Writer	The teacher thought actively about the writer behind the paper, not just noting what the person did wrong or does not know, but considering what the writer thought, wanted to achieve, was trying to accomplish, decided, and so on.
Reader Response	Consistent with the work of Louise Phelps (1989), the teacher approached student writing from the perspective of a reader, noting what makes sense, is easy to follow, is clear, is convincing, and so on, rather than as an evaluator determine how right or wrong the writing is. See Brian Huot's (2002) excellent chapter, "Reading Like a Teacher" pp. 109-36.
Positive Focus	Rather than commenting first, and extensively, about the faults of the paper, the teacher made a balanced analysis of the paper that points out strengths as well as faults. This is helpful to students in better understanding the characteristics of their writing, as well as to make writing a more rewarding experience.
Five-Paragraph Essay Challenges	The teacher raised a question or challenge to the college faculty's derogatory view of the five-paragraph essay. These comments were often followed by explanations of why the form is useful in teaching high school writing.
Conventions Questioning	The teacher raised a question about the value of teaching the conventions of written English (spelling, punctuation, and so on) relative to the value of teaching higher-order concerns such as the quality of the ideas or the argument. Their concerns related both to how the conventions should be weighted in grading and at what grade levels they should be emphasized.
Stress of Teaching	The teacher expressed frustration over conflicting expectations or pressures in their teaching jobs. Most of the comments were about the pressure to prepare students for the state tests (the ACT) while at the same time not being allowed to fail students who do not complete assignments.
Expressions of Disagreement	The teacher expressed an opinion contrary to the opinion of someone else, or point out an error of fact in another person's post.

Perspective	Definition
Introspective Comments	The teacher voiced inner thoughts about the scoring/grading process, his or her own standards, values, or beliefs compared with other teachers or with the college faculty.
Expressions of Agreement	The teacher expressed agreement with another teacher's comments, and sometimes that also recognized the other person's contributions or offered praise or moral support.
Questions to Group	The teacher raised a question for other teachers to answer, sometimes a factual matter, but most often an opinion. Many of the questions spurred discussion and exchange of ideas.

Appendix L: Notes about Technical and Practical Issues

Technical Considerations for Future Applications

Although the technical features of the Web site were not directly under study, I identified a few items that should be mentioned for future researchers.

First, the Web site was constructed on Michigan State University's course management program, Angel. Because of this, a few more clicks were required for logging onto the site than if it had been built directly on the Internet, but that did not seem to concern anyone. One person had trouble logging on to the first time, but quickly learned how to do it. An ordinary Web site developed outside Angel would probably facilitate the logging on procedure and possibly allow other functions to be programmed with more specific affordances and limitations.

Downloading the PowerPoint slides was slower than people liked, and perhaps that could be speeded up with different technologies. High-speed connections were definitely required for the PowerPoints. In future research, I hope people more sophisticated than I could improve the ease with which the recordings were made, transported over the Internet, and edited. Audacity is a remarkable free program and easy to learn, but possibly another option would be more serviceable for the size and type of files this program requires.

Creating the PowerPoint slides was time consuming largely because it was a new activity to me. If I were to do this in the future, I would rehearse the college faculty in how to prepare their commentaries and record them so they would require less editing. This would entail developing an outline in advance of recording and deciding which part would be described by which person, and testing sound levels with the microphone.

When they did these things at the end of the project, editing was fairly straightforward. A better microphone that masked background noise would be helpful. Also, I would try to solve the only real technical problem we encountered: the audiofiles, which had been created in Audacity, were too large to email from Blue Water College, so they had to be retrieved in person.

Recording the Audacity files required the college teachers to meet in pairs during the summer when the building that housed their department was locked; the microphone was of poor quality; sound levels were difficult to establish, the mike picked up background noise, sometimes the sound was scratchy; the Audacity files were too large to email; and just learning to use Audacity was challenging for some. One of them found the first Round especially nerve-wracking, and suggested having a dry run before another project like this actually starts. Technical assistance on this aspect of the project would be a wonderful timesaver and almost certainly result in superior products.

Practical considerations for future implementations

I learned a few things from this study that did not necessarily appear in the data, but should be passed along for future researchers. First, assessment communities should probably be instituted during the summer when teachers of high school English and faculty of first-year college composition courses are not already burdened by piles of their own students' papers to grade. Even with attractive incentives, teachers are not likely to volunteer for this kind of program during the school year.

Second, the audio recordings made by the college teachers were a key element. Their voices brought their comments to life and personalized them in a way that could never be achieved by printed text alone. Also, having a pair of voices that exchanged

ideas, usually reinforcing each other but occasionally expressing different views, seems essential for preventing the "expert" authority from coming across as dictatorial and impersonal. In a few instances where the pairs recorded separately, one after the other, the effect is not nearly as convincing and appealing as a more collaborative dialog about the papers; and moreover, recordings that interspersed comments of both colleagues in a conversational fashion modeled the collegial and respectful conversations that became the norm in the study. Listeners were charmed and engaged, I believed, when the college commenters supported each other, added to each others' points, and even when they disagreed in a familiar and collegial manner.

As an example of how powerfully personalities either "reach out to" or "turn off" listeners of the recordings, one of the recording pairs early in the project divided their commentary into positive and negative segments, with one person reporting on the favorable aspects of the paper and the other reporting the negative. Both teachers had otherwise likeable and appealing voices on the tape, but this effort was not successful. It produced a "good cop, bad cop" effect, and I, along with the high school teachers, took an immediate dislike to the "bad cop."

Also, some rehearsing and other preparation for the recordings is helpful. On several occasions, it seemed that the commentary strayed from main points summarized early in the recording, and in one case even contradicted them. A clear delineation of points to be made and an outline for organizing them would help develop coherence throughout the slide show for a single paper, and would certainly reduce the amount of time needed for editing.

Also, part of the preparation should be, whenever possible, referring to specific phrases, passages, or paragraphs of the paper to use as examples. A tremendous difference resulted from a statement such as, "The paper contains redundancies," and a statement such as, "Notice the redundancy in paragraph 4, where the writer says..." Examples that illustrate the points of analysis should be quoted on every slide except for those that introduce or summarize points. It is grounding in actual examples of student-composed text that makes the ideas of the expert group meaningful to others.

Finally, I believe the Web-based version of the assessment community is the key to making it accessible, convenient, and usable. High-speed Internet access is still not available in some areas of the country, but where it is available, the assessment community can function without tiring meetings that require rigidity in people's calendars. Two of the participants in this project worked early in the morning, and at least one worked late at night. One usually went online and completed the bulk of her work in one large chunk on Sunday evenings; several others worked in short bursts during the week. All but one of them could work at home, interspersing their posts with family activities, laundry, cooking, and child care. The home-based nature of the study was especially important in this region, in which schools and homes were spread over a five-county area and drive time to meetings would have amounted to an hour and half per meeting. It may be equally important in urban areas where drivers fight traffic congestion or use public transportation. Adding energy and climate considerations to the equation, online communities make great sense if they work, and the results of this study indicate that is possible. Perhaps the greatest argument for online communities is their potential

for linking people across great distances and time zones, and more importantly, across divides of class, race, ethnicity, and other cultural boundaries.

REFERENCES

References

AERA, APA, & NCME. (1999). *Standards for Educational & Psychological Testing.* Washington, DC: AERA.

Akkerman, S. et al. (2007). Reconsidering group cognition: From conceptual confusion to a boundary area between cognitive and socio-cultural perspectives? *Educational Research Review, 2,* 39–63.

Barab, S., et al. (2006). Our designs and the social agendas they carry. *Journal of the Learning Sciences, 16(2).* 263-305.

Belanoff, P. (1991). The myths of assessment. *Journal of Basic Writing, 10.* pp. 54-56.

Bielaczyk, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *Journal of the Learning Sciences,* 15(3). 301-329.

Broad, B. (2003). *What we really value.* Logan, UT: Utah State University Press.

Bruffee, K. A. (1999). *Collaborative learning: Higher education, interdependence, and the authority of knowledge.* Baltimore, MD: The Johns Hopkins University Press.

Carey, K. (2009). Blocking public comparisons obstructs knowledge too. *The Chronicle of Higher Education.* Feb. 20. Retrieved Feb. 20 from http://chronicle.com/weekly/v55/i24/24a07201.htm

Chait, R., & Venezia, A. (2009). Improving academic preparation for College: What we know and how state and federal policy can help. Washington, D.C.: Center for American Progress.

Chappell, V.A. (1991). Teaching like a reader instead of reading like a teacher. In V.A. Chappell, M.L. Buely-Meissner, & C. Anderson, (Eds.), *Balancing acts: Essays on the teaching of writing in honor of William F. Irmscher.* Carbondale, IL: SIU Press.

Coe, R. M. (1992). Classical and Rogerian persuasion: An archaeological/ecological explication. In *Rogerian perspectives: Collaborative rhetoric for oral and written communication,* ed. Nathaniel Teich, 83-100. Northwood, N.J.: Ablex.

Condon, W. (1997). Building bridges, closing gaps. In Yancey, K.B., & I. Weiser, Eds., *Situating Portfolios: Four perspectives.* Logan, UT: Utah State University Press. 196-213.

Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* Thousand Oaks, CA: Sage.

Cronbach, L. (1989). Construct validity after thirty years. In R.L. Linn, (Ed.), *Intelligence: Measurement Theory and Public Policy.* Urbana, IL: University of Illinois Press. 147-71.

Culham, R. (2003). *Six + one traits of writing: A complete guide (grades 3 and up).* Portland, OR: Northwest Regional Educational Laboratory.

Derrida, J. (1976). *Of grammatology.* Spivak, G.C., Trans. Baltimore: The Johns Hopkins University Press.

Elliot, N. (2005). *On a scale: A social history of writing assessment in America.* New York: Peter Lang.

Faigley, L. (1989). Judging writers, judging selves. *College composition and communication, 40,* 395-412.

Fischer, G., & Giaccardi, E. (2004). Meta-design: A framework for the future of end-user development. In Lieberman, H., Paternò, F., & Wulf, V. (Eds.), *End user development: Empowering people to flexibly employ advanced information and communication technology.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Fish, S.E. (1980). *Is there a text in this class?: The authority of interpretative communities.* Cambridge, MA: Harvard University Press.

Fontana, A. & Frey, J.H. (2005). The interview. In N.K. Denzin & Y.S. Lincoln, (Eds.), *The Sage handbook of qualitative research,* 3rd. ed. Thousand Oaks, CA: Sage.

Foucault, M. (1970). *The order of things: An archeaology of the human sciences.* New York: Pantheon.

Garson, G.D. (2008a). Statnotes: How should a negative reliability coefficient be interpreted? Retrieved February 19, 2009 from http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm

Garson, G.D. (2008b). Statnotes: Topics in multivariate analysis. Retrieved February 15, 2009 from http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine/Atherton.

Harrington, S., Rhodes, K., Fischer, R.O., & Malenczyk, R. (2005). *The outcomes book: Debate and consensus after the WPA Outcomes Statement.* Logan, UT: Utah State University Press.

Huot, B. (2002). *(Re)articulating writing assessment for teaching and learning.* Logan, UT: Utah State University Press.

Kirst, M.W. & Venezia, A., Eds. (2004). *From high school to college: Improving opportunities for success in postsecondary education.* San Francisco: Jossey-Bass.

Lave, J., and Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* New York: Cambridge University Press.

Lederman, D. (2009). Whither accreditation? *Inside Higher Education.* Jan. 28. Retrieved Jan. 28 from http://www.insidehighered.com/news/2009/01/28/accredit

Legg, S.M. (1998). Reliability and validity. In W. Wolcott & S.M. Legg, Eds., *An overview of writing assessment: Theory, Research, and Practice.* Urbana, IL: National Council of Teachers of English.

Lumina Foundation for Education (2009). A stronger nation through higher education: How and why Americans must meet a "big goal" for college attainment. Indianapolis, IN.

Messick, S. (1993). Validity. In R. L. Linn (Ed.), *Educational measurement*, 3rd ed., 13-103. Phoenix, AZ: Oryx Press.

Miles, M.B. & Huberman, A.M. (1994). *Qualitative data analysis.* 2nd ed. Thousand Oaks, CA: Sage.

Mintzberg, H. (2000). The professional bureaucracy. In M.C. Brown, (Ed.), *Organization and governance in higher education* (5th ed.) (pp. 50-70). Boston: Pearson.

Morgan, G. (1997). *Images of organizations, 2nd ed.* Thousand Oaks, CA: Sage.

Nardi, B.A. (2005). Placeless organizations: Collaborating for transformation. *Mind, Culture, and Activity, 14(1-2).* 5-22.

National Commission on Excellence in Education of the U.S. Department of Education. 1983. *A nation at risk: The imperative for educational reform.* Washington, D.C.: Author.

Paavola, S., Lipponen, L., & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research, 74(4)*, 557–576.

Palomba, C.A., & Banta, T.W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education.* San Francisco: Jossey-Bass.

Sanoff, A. (2006, March 10). Companion survey of high school teachers and college faculty on students' preparation for college. Results reported in A perception gap over students' preparation. *The Chronicle of Higher Education, 52* (27), B9.

Spellings Commission on the Future of Higher Education of the U.S. Department of Education. (2006). *A test of leadership: Charting the future of higher education.* Washington, D.C.: Author.

Stahl, G., (2006). *Group cognition: Computer support for building collaborative knowledge.* Cambridge, MA: MIT Press.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory procedures and techniques.* Thousand Oaks, CA: Sage.

Valid Assessment of Learning in Undergraduate Education (2009). Retrieved February 21, 2009, from http://www.aacu.org/value/

Voluntary System of Accountability (2009). Retrieved February 21, 2009, from http://www.voluntarysystem.org/docs/cp/LearningOutcomesInfo.pdf

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* New York: Cambridge University Press.

White, E.M. (1994). *Teaching and assessing writing,* 2nd ed. Portland, MA: Calendar Island Publishers.

Whitehouse.gov. (2009). Retrieved March 22 from: http://www.whitehouse.gov/agenda/education/

Yancey, K.B., & Weiser, I. (1997). *Situating portfolios: Four perspectives.* Logan, UT: Utah State University Press.