2

ASSESSMENT IN A CULTURE OF INQUIRY
THE STORY OF THE NATIONAL WRITING PROJECT'S ANALYTIC WRITING CONTINUUM

Sherry Seale Swain

National Writing Project

Paul Le Mahieu

Carnegie Foundation for the Advancement of Teaching

In order for assessment to support student learning, it must include teachers in all stages of the process and be embedded in curriculum and teaching activities. It must be aimed primarily at supporting more informed and student-centered teaching. . . . Like students, teachers also learn by constructing knowledge based on their experiences, conceptions, and opportunities for first-hand inquiry.
    Linda Darling-Hammond (1994, p. 25)

Every time I come to [an AWC scoring conference] and we do the practice papers and the anchor papers, it's good to hear everybody's input about why they score things a certain way. . . . Now as a table leader, I can see the scores of others, and I say . . . "I'm interested in your thinking behind this." So that just adds another dimension to it for me, to see what it is that people see as a [score of] 3 in *Stance* because I see it as a [score of] 2. It's just another layer of it for me.
    Teacher

Teachers, the heart of education, too often find themselves excluded from the very process that policymakers advise should be central to instruction, namely assessment. Across this country, K–12 faculty groups gather regular-

ly for sessions on "data driven" instruction. They pour through pages of statistical materials, and then they go back to their classrooms and do what they've always done: the very best they can with the information they have. What's missing in this picture? We suggest that what's missing is the teacher's involvement in assessment in any meaningful way: helping to design it, learning from it, and using it to improve instruction.

This is the story of how teachers thinking together with writing assessment experts helped to create a technically sound and rigorous writing assessment, one that is useful in the classroom as well as in research. The system diminishes the conundrum described by White: that often teachers feel "forced to choose between tailoring their teaching to an impromptu test and helping their students learn how to write . . ." (2007, p. iv). At the center of this story is a cohesive educational community, imbued with a vital inquiry stance, that developed, investigated, refined, and expanded the uses of the assessment system over an extended period of time. The Analytic Writing Continuum (AWC), developed by the National Writing Project (NWP), offers an opportunity to explore the potential of assessment that is locally contextualized yet linked to a common national framework and standards of performance.

The National Writing Project, arguably the nation's most long-standing educational reform effort, with its history of respect for teacher knowledge and reflective practice, serves teachers at all levels, from early childhood through university. NWP provides professional development and resources for teaching and learning, improving the teaching of writing, and conducting research on the teaching and learning of writing in classrooms and schools. In this instance, NWP addressed the issue of writing assessment guided by the same principles through which it has traditionally addressed writing, learning, and teaching: first by calling on the expertise of practitioners—teachers and researchers in the field; and second by inviting adaptations, inquiry, and feedback from those who put the assessment to work—in scoring events, classrooms, and professional development. These principles provided the basis for decisions during the development of the system and the refinements that continue to press upon the system, keeping it relevant and useful for its various purposes.

THE TIME WAS RIGHT

The AWC system addressed a specific need for writing assessment within the NWP network that emerged when, in the fall of 2003, a cohort of six research groups from writing project sites across the country received grants from NWP to study the effects of their programs on instructional practices

and student writing performance. In addition to collecting descriptive and demographic data, each of the research sites committed to include a comparative reference in order to compare writing growth of students whose teachers were involved in writing project professional development with that of students whose teachers were not yet involved with a writing project. Further, the grants stipulated that growth in writing be measured by a direct assessment of writing. For the most part, local groups lacked the financial resources and assessment expertise to develop rigorous scoring systems themselves. Credibility was an even bigger issue. How could a local site design and deliver the program, develop, administer, and score the assessment, and then expect others to respect the outcomes they reported? NWP recognized the need for impartial judgments of students' writing achievement independent of these local sites. In 2004, with the second cohort of research sites, NWP committed to providing a national scoring system that would operate independently of the sites and provide unbiased and credible scores.

NWP researchers needed a robust assessment system that would serve not a central research design administered in multiple locales but rather a number of locally designed research studies, each uniquely suited to its context. Although local research teams would be committed to the direct assessment of student writing that included data from both program and comparison groups, there would be differences in research designs in terms of prompts, administration procedures, research questions, and analyses. This single system would require the strength of common standards and procedures and the flexibility to accommodate multiple prompts, multiple genres, multiple grade levels, and varied conditions for writing—from first draft on-demand writing to fully revised pieces from student portfolios. In short, NWP needed a rigorous central system that could speak to local questions and needs.

Initially, a group of NWP researchers—classroom teachers and university directors—met to determine whether an assessment system that would meet their needs might already exist or whether an existing system might serve as the basis for a new one. Over a 2-day period, practitioners articulated their beliefs about good writing assessment. "The substance of the writing must outweigh emphasis on conventions." "The focus must be on defining the quality of writing." "We need to be able to accommodate the grade levels and prompts that each study generates." "We need to be able to see growth where there is growth." The group outlined the requirements of such a system as follows:

- *A focus on the attributes of writing that the NWP teachers, programs, and researchers valued;*
- *A focus on the quality of the writing itself;*

- *Potential to address multiple grade levels, genres, and prompts;* and
- *A scale sensitive enough to detect differences for research purposes.*

After reviewing scoring systems from a number of English-speaking countries, the group came to consensus that the Six + 1 Trait Writing Model (Culham, 2003) came closest to their requirements. As an added advantage, the Six + 1 Writing Trait Model had considerable currency among practitioners and policymakers: most teachers of writing were familiar with the system, and a number of local and state education agencies had already based their writing assessments on it. The group decided that the NWP assessment system would have its roots in this well-known and widely used system.

CREATING THE AWC SYSTEM

In 2004, following the recommendation of the initial group of practitioners, a national panel of experts on student writing, along with senior NWP researchers,[1] confirmed the choice of the Six + 1 Trait Writing Model as the starting point for the new system. This group began their task by reviewing the model to determine in what areas it was and was not suitable for the research purposes of the NWP sites. For example, much of the language, although totally appropriate for stimulating conversation between teachers and students, was not sufficiently precise or rigorous for use in a research situation. The group decided to rethink the focus on "the writer" and "the reader," which makes scoring difficult in a situation in which neither the writer nor the intended reader is known to the scorer. Additionally, some definitions of the traits caused concern, for example, references to the writer's "personal details" as a requisite for demonstrating strong voice. With permission and encouragement from the originators of the Six + 1 Trait Writing Model, the panel set about making the following modifications to create what is now known as the NWP Analytic Writing Continuum Assessment System.

Conceptual Coherence

Revisions to particular traits of the Six+1 Trait Writing Model brought about conceptual coherence to the AWC Assessment System, enhancing reliability and validity and framing attributes for research and classroom uses.

Content. Whereas the Six+1 Trait Model was concerned with Ideas/Content that focused on the use and shaping of details, the AWC expansion of Content (including quality and clarity of ideas and meaning) includes language to address a variety of genres with suggestions for a multitude of types of support appropriate to academic as well as narrative writing. At score point 6, one element of the Content attribute describes the writing as containing "ideas that consistently and fully support and/or enhance the central theme or topic (e.g., well-developed details, reasons, examples, evidence, anecdotes, events, and/or descriptions, etc.)."

Structure. Reconceptualizing the Organization trait, which focused on a seamless overall integration, led to the AWC Structure attribute, including internal as well as overall attention to coherence and unity. Structure also addresses formulaic organization. At score point 3, one bulleted thread reads, "Includes a structure that is formulaic and predictable, or occasionally erratic, inconsistent, or uneven." At score point 6, that same thread reads, "Presents a compelling order and structure; writing flows smoothly so that organizational patterns are seamless."

Stance. Reconceptualizing the Voice trait led to the newly defined attribute of Stance—the presence of a clear and appropriate perspective (Dipardo, Storms, & Selland, 2011). Whereas people often interpret voice as a reflection of the writer's personality and then construe it as a "perky" or "excited" tone, Stance is concerned with the writing itself: its perspective, tone and style, purpose and audience, and level of formality. A scientific laboratory report should not be perky or pensive, but rather authoritative in its account of procedures, results, and interpretations. Descriptors of the Stance attribute at score point 6 are shown in Figure 2.1.

---

6. The writing:

- Consistently and powerfully demonstrates a clear perspective through tone and style.
- Consistently demonstrates a distinctive and sophisticated tone or style that adds interest and is appropriate for purpose and audience.
- Exhibits level(s) of formality or informality very well suited for purpose and audience.

Fig. 2.1. NWP Analytic Writing Continuum. Descriptors of the Stance Attribute at Score Point 6.

Sentence Fluency. The AWC values logical, clear sentences that exhibit appropriate rhythm, flow, and variation in structure and length. Intentional and/or effective use of fragments is noted in score points 4, 5, and 6.

Diction. The term *diction* signals appropriate attention given not just to words but also to expressions and phrasing. The Diction attribute also addresses appropriate and inappropriate modifiers as well as strong nouns, lively verbs, imagery, and metaphor, as appropriate to audience and purpose.

Conventions. In the AWC this includes usage, spelling, punctuation, and capitalization, and is defined at each score point by the ease or difficulty of reading and the amount of editing that would be required for publishing, from "almost no editing" for score point 6 to "extensive editing" for score point 1.

Scoring Scale

Extending the scale from four to six points gave the scoring system increased sensitivity to differences among pieces of writing, allowing the detection of change over time or differences between groups of students. In addition, a new "look" for the AWC reflects the NWP belief that, in reality, writing scores fall along a potentially infinite continuum of score points—even though only six of those points are available as actual scores.

Focus

Focusing the evaluative judgments exclusively upon the student writing centered the assessment on writing rather than students. Thus, instead of the stem for each score point beginning with "the writer," the focus in the AWC is on "the writing" itself. In deference to emerging writing skills, the panel utilized language at each score point to describe "what is there" rather than deficiencies or "what is not there," so that a determination of where the writing falls along the continuum reveals direction for improvement. Figure 2.2 shows the Diction attribute at score point 3, describing what is present in the writing rather than what is lacking.

Language

Rather than "traits," the AWC system addresses the six attributes described earlier with identifying elements or "threads" within each that more clearly define the attributes, clarify the differences among score points, and enhance scoring reliabilities. Three to five threads, denoted by bullet points, run

3. The writing:

- Contains words and expressions that are sometimes clear and precise.
- Contains words that are primarily simple and general, yet adequate.
- Contains mostly bland verbs or commonplace nouns and inappropriate modifiers.
- May include imagery or figurative language; when present it is simple and generally not effective.

Fig. 2.2. NWP Analytic Writing Continuum. Diction Attribute at Score Point 3.

across the score point descriptors and define the differences among scores. For example, the first thread in the content attribute describes the degree to which a central focus is shaped and presented, beginning with "clear and consistently focused; exceptionally well shaped and connected" at score point 6; and at score point 2, "May present several ideas, but no central focus emerges; seldom shaped and connected."

Holistic and Analytic Scoring[2]

In an effort to provide data that would serve each of the two primary purposes of the assessment system (i.e., evaluating NWP programming and providing data to teachers to inform instruction), the AWC applies both holistic and analytic scoring procedures. The holistic scoring guide describes the elemental components of the holistic score in the same way as the analytic framework so that the training for the analytic attributes also deepens understanding of the holistic scoring. However, whereas some systems build holistic scores as arithmetical aggregates of a set of analytic scores, the AWC does not attempt to do so. It preserves the analytic scores as separate scores, allowing both evaluative investigation of programs in specific areas of writing performance as well as diagnostic and instructionally relevant data. The holistic score is a single summary judgment about the quality of the writing. Research conducted within the AWC community affirms that the procedure of scoring holistically prior to analytic evaluation preserves the holistic score as independent of those assigned in analytic scoring (Singer & LeMahieu, 2012). In this way the holistic score can be regarded as "other than the sum of some identified parts" (a phrase chosen advisedly as, in fact, the complex

and interdependent nature of writing performance means that the whole may be more or less than its deconstructed and abstracted components).

Interestingly, because it is handled in this way, the AWC system implicitly addresses the three "major problems" that White identifies as restricting the utility of analytic scoring (White, 1994):

- Lack of agreement about the subskills of writing: Notwithstanding the need for a thorough empirical investigation of the structure of the framework, the AWC (which is ongoing within the AWC community) is built upon an examination of major assessment systems and their descriptive frameworks. The consistencies in these frameworks and the widespread applicability of the AWC across policy and instructional settings strongly suggest that it addresses just such an implicit consensus.
- Difficulty of obtaining reliable subscores: The AWC system (with its anchors, training, and calibration samples, as well as training, calibration, scoring, and performance monitoring processes) has achieved levels of reliability that are quite exemplary and certainly adequate to its purposes.
- Time consuming and costly: Obviously scoring analytically and holistically takes longer than holistic scoring alone, yet the expanded information that the AWC system yields has been demonstrated to be of value to and appreciated by researchers and teachers. (Swain, LeMahieu, Sperling, Murphy, Fessehaie, & Smith, 2010)

Determining What "Good" Looks Like

The AWC Range Finding Team meets annually to review the anchor, calibration, and training papers that set and convey standards for each of the levels of scoring: lower elementary (grades 3 and 4), upper elementary (grade 5), middle school (grades 6–8), and high school. For each scoring level, these sets include six anchor papers, with similar scores for each of the attributes; two sets of six practice papers, illustrating mixed scores among the attributes; and a growing body of recalibration papers. These sets explicate the genre appropriate to the AWC: informative, persuasive, and narrative. Initially, the team reviewed available papers for each of the scoring levels and established the anchor sets. These remain largely unchanged in order to maintain standards across time and place. Annually, however, "new" papers from the current set to be scored replace two or three papers in the two practice sets to reflect current prompts and to focus on emerging issues such as economic, cultural, and language diversity among student writers.

Entering the Inquiry Community: Scorers and Learning to Score

Entry into the AWC community begins with the selection of scorers, almost all of whom are NWP Teacher-Consultants (TCs). Local site leaders nominate teachers who have participated in the 5-week invitational summer institute and who have strong potential as scorers. These leaders see the involvement of TCs as opportunities to build local capacity for serving school districts. During the 6 to 8 hours of training and preparation in the AWC system, scorers become familiar with the attributes, and calibrated to the standards. Room leaders and table leaders come from the range finding team and from the pool of experienced teacher-scorers who have shown expertise in the system and who have enhanced that expertise by using it in their schools or writing projects.

During interactive reading and discussion of the system, teacher-scorers encounter the tenets underlying the Analytic Writing Continuum—for example, that the continuum actually describes a graduated range of performance. The scorer's task is to determine which of the available score points most accurately describes each attribute of the writing. The interaction at the training looks like this: The room leader asks the group of scorers, "Why does this anchor paper represent a score of four in Content?" The teachers begin a process of conferring between the paper at hand and the descriptors of the AWC. They ask questions—of the room leaders, of each other, of themselves. The discussion helps the scorers to embrace the AWC, to reflect on it, and to help refine it. Rather than being limited to a narrow definition at each score point, scorers learn to recognize a range of characteristics and writing abilities clustered around each score point. As important, TCs examine writing in a new way—with a shared sense of quality, just what "good" is and what it looks like in its many expressions in their classes. The result is a commonly held perspective on and expectation for quality and one that is typically higher than any one individual arrived with.

The training also includes mini-lessons illustrating several of the attributes. Scorers participate in a lesson that uses photos to compare writing structures to various architectural structures, the predictable tract house to a formulaic rendition of a piece of writing: "Does this house remind you of a formula with its box shape, symmetrical windows, unremarkable entry, and predictable landscaping?" "Does this building that houses a modern art museum appear constructed for its specific purpose?" "How does the structure of this paper take readers on a journey that shifts forward and backward in time; how does its structure support its purpose?" Through another mini-lesson, scorers discover a seemingly infinite number of stances, from doubtful to persuasive, from humorous to mournful. In the course of the PowerPoint presentation, scorers note postures, tones, styles, and ultimately, the presence of appropriate stance in a piece of writing.

Prompts and Papers

Scorers encounter a wide variety of prompts at each scoring conference. All the local studies employ quasi-experimental designs that include direct assessment of writing, comparative references (usually a comparison group of students whose teachers are not involved in writing project professional development), counterbalanced prompts to account for an effect in which one or another of the prompts might be advantageous to a group of students, and pre/post administration of prompts so that differences over time can be observed. The result is quite a complex array of categories into which student papers fall at the scoring conferences. For example, papers from a single research study focused on a single grade level would fall into eight groups based on the categories of pretest/posttest, prompt A or B, and group.

Add to this mix research designs that encompass additional grade levels or research questions dealing with a variety of genre and the complexity expands. Across the multiple research designs, complexity increases in other ways as well, with variation in such elements as the following:

- Prompts. Local research teams select or design their prompts to reflect foci of their studies, from writing across the curriculum to descriptive, persuasive, or informative writing. Further, because all the research studies employ counterbalanced designs, there are at least two prompts for each study; for example, 10 studies contributing papers would mean a minimum of 20 prompts in play at any one scoring conference. With more complex research designs come even more prompts, sometimes including portfolio pieces that have undergone revision.
- Administration of prompts. Again, local research teams design the administration of the prompts, often mirroring state testing practices regarding timed or untimed administration. Because of local regulations related to classroom access, some sites send written instructions to the schools or teachers regarding test administration. In other localities, writing project leaders go into the schools to train the teachers to administer the assessments; in still others, specially trained writing project teachers administer the prompts. In each case, however, the same conditions prevail across program and comparison groups.
- Paper preparation. Local sites remove identifying information so that scorers do not know any specifics of the writing sample being evaluated (e.g., identity or even gender or race of the student, place of origin, group [program or comparison], or time of administration [pretest or posttest]). Local research teams code the papers and keep complete databases for purposes of analysis

and reporting. In most cases, papers and their scores are eventually returned to the schools and teachers.

- Statistical analysis. NWP provides individual student scores for the holistic and the six analytic attributes to the local research teams, each of which employs a statistician to conduct the final analyses for the study. From these analyses, the local research teams write their reports. Researchers at NWP then compile these findings into reports and research briefs (NWP Research and Evaluation Unit, 2010).

Scoring Procedures

Scorers calibrate to a criterion level of performance during the initial training, and then recalibrate following every major break in the scoring (meals and overnight). Fifteen to twenty percent of the student writing is scored twice, in part so that reliabilities can be calculated and in part so that the performance of the scorers can be monitored. Room and table leaders then provide individual recalibration sessions as needed. Over the six scoring conferences, reliabilities (measured as interrater agreement, defining agreement as two scores being identical or within one single score point of each other along the continuum) ranged from 89% to 93%, with an aggregate across all scores of 90% (see Table 2.1). In addition, at each scoring event, a small number of papers or scorers might follow special procedures to allow the investigation of some aspect of writing assessment that is of common interest (e.g., altering the order of scoring the attributes in the analytic scoring to explore whether order influences the scores). In these instances, the scores are not treated as operational because researchers prefer that all operational scores are obtained from the established procedures. Finally, NWP researchers infuse a constant set of papers (an equating set) taken from the first scoring conference into the papers to be scored at each subsequent scoring event in order to monitor the consistency of standards as they are applied across years.

To date, six national scoring conferences, with the Analytic Writing Continuum as the centerpiece, have produced student outcome data for 18 local site studies, spanning grades 3–12, all using quasi-experimental designs (NWP Research and Evaluation Unit, 2010). Over the 6 years, 397 TCs from 50 local NWP sites in 19 states have calibrated as scorers, including 65 experienced scorers who serve as room and table leaders. Every scoring conference includes time for reflection among the teacher-scorers. These reflections focus on both the AWC itself as well as its implications for the teachers' instructional practices. These reflective comments and conversations form the basis for continuous refinements to the system, yet the primary purpose of the NWP Scoring Conference is just that—obtaining valid and reliable scores to be used across multiple research studies.

TECHNICAL RIGOR AND QUALITY

As with many high-quality assessment systems, developers attended to technical rigor and quality from the outset. What sets the AWC apart from most, however, was that the inquiry-oriented stance of the NWP encouraged an improvement-oriented approach to its development. In short, technical quality was not a "state of grace" demonstrated at a singular point in time but a necessary aspect of the life and development of the system.

Reliability

Constant examination of scoring reliabilities, including the daily reliability reports that scorers eagerly awaited, ensured adequacy of the reliabilities. Leaders and scorers paid particular attention to aspects of the system that were undergoing revision at any point in time as well as those that consistently posed challenges. Table 2.1 presents the operational reliabilities over the life of the AWC. It shows that the interrater agreements ranged from 83 to 95%, certainly adequate to the research that was the first purpose for the development of the system. Although there are many observations that will occur to the reader, these reliabilities reveal at least two noteworthy points: first, they suggest areas that proved more challenging for scorers to understand and apply (e.g., Stance, which was addressed by modifications to the assessment framework itself as well as the training procedures); and second, they reveal a general trend of improvement over time—indicative of the impact of the research-driven modifications and the increased understanding of the scorers.

Validity

Examining validity in multiple ways adds to the credibility of the system. To offer just a few examples: Construct validity is a constant focus of attention through factor and facets analyses that support the fundamental conceptual structure of the system. Correlating scores from the AWC with other measures, notably state writing assessment systems, allows researchers to explore concurrent validity. Several local research teams have successfully included these analyses in their research designs. These analyses demonstrate appropriately high correlations and interestingly suggest that the AWC provides higher standards of performance and more useful information for instruction. Finally, the important issue of consequential validity is an ongoing focus of considerable effort and inquiry as reported in the next section of this chapter.

Table 2.1. National Writing Project Analytic Writing Continuum Reliabilities by Year of Scoring, 2005–2010

Year	Total number of Papers	Number of papers double scored	Total adjudi- cations	Agreement							
				Overall	Holistic	Content	Structure	Stance	Sentence Fluency	Diction	Convention
2005	7505	1748	1004	92%	95%	92%	91%	90%	91%	93%	91%
2006	6493	1071	661	91%	94%	92%	90%	88%	90%	93%	91%
2007	5362	1120	948	88%	90%	89%	89%	85%	86%	90%	86%
2008	6549	1287	1200	87%	90%	88%	87%	83%	86%	88%	86%
2009	6763	1397	969	90%	93%	91%	91%	86%	89%	90%	90%
2010	4827	1563	788	93%	94%	94%	94%	91%	92%	93%	91%
2005 –2010	37,499	8186	5570	90%	93%	91%	91%	89%	89%	91%	89%

Note: The Overall reliability for each scoring conference was computed using the total numbers of scores and adjudications.

Utility of Use

A final area of inquiry monitors the system in the interest of ensuring its utility to various users. Ongoing equating studies in which a large set of student papers is scored across years and settings ensure the comparability and consistency of standards. Newly initiated vertical equating studies in which sets of papers are scored at multiple grade levels monitor the interpretation of the standards at each level of the system and interest those concerned with writing growth over time.

ADDING TO THE KNOWLEDGE BASE ABOUT WRITING ASSESSMENT

Two factors allow researchers to use the AWC to explore larger assessment questions. The first is its longevity: Assessment systems don't often remain stable long enough for researchers to use them as a tool for exploring assessment in and of itself. Second, the community of use surrounding the AWC welcomes inquiry and professional reflection. In the first 3 years of its implementation, researchers used the scoring conference to explore a number of research questions: When involved in both holistic and analytic scoring, does it matter if scorers focus first on the holistic or analytic scores (Singer & LeMahieu, 2012)? Does the order in which scorers are asked to assess the six attributes of writing influence the scores assigned to those attributes? How do scorers conceptualize and regard "voice," and what elements and/or language influence scorers' evaluative judgments about it? (DiPardo et al., 2011). More recently two research teams conducted focus groups, one to examine the implications of prompt design and the other looking at the characteristics of English learner writing that might position us to better evaluate it in an inclusive system of assessment.

Impact on Scorers

The Scoring Impact Study (Swain et al., 2010), launched in 2008, determined the extent to which the NWP's assessment system has any value to those who serve as scorers. Initially 325 teacher scorers received invitations to complete an online survey. One hundred thirty-six (41.6%) responded. Data for the study included the online survey, focus groups with 20 respondents, and interviews with 14 scorers. It revealed how teachers' experiences with the AWC had influenced their beliefs about writing and assessment, their

classroom practices, the nature of their local writing project work, and the nature of writing itself. Influences of the AWC scoring experiences on teachers fall into four categories:

1. Understanding the Characteristics of Writing. The AWC experience is a catalyst for deeper understandings of three elements: characteristics of writing, assessing student writing, and teaching students to write. On a 6-point scale, 52% to 69% of responding teachers marked the highest rankings (5 or 6), indicating that the assessment system becomes a centerpiece for inquiry, as teachers embrace it and use it as a lens to question their practice. One respondent contemplated her newfound insights on teaching sentence fluency:

    > It had not occurred to me to actually teach sentence fluency. I guess I thought that it could somehow be a natural outflow of reading, and yet after the conference, I realized that I could actually teach that actively, rather than hope they would pick it up. I could do it more intentionally. And so as a result of the conference, I just started really looking at teaching, not just sentence fluency, but all of the six attributes in a more intentional way.

2. Changes in Practice. Teachers take the AWC into their classrooms as a tool for evaluating student writing, as a guide for discussing writing with students, and as a tool for students to use to examine and ultimately improve their writing. Again, survey respondents marked top-ranking choices over half the time, 56%, 67%, and 64%, respectively. Fifty percent indicated the experience had changed their views about what students should learn. One teacher described varied practices for helping young writers to structure their writing:

    > I write a piece in two different ways, one with a very weak or no opening and closing and another with strong ones. I let students pick which one they like and then tell why. . . . I read them lots of different openings and say, "Now what was this?" And they'll say, "It was a question or it was a shocking statement or an amazing fact." . . . We make a chart of choices for beginnings and endings. If they are writing an information piece, I might offer Laurence Pringle books because he has strong openings, and strong audience awareness. He starts his bat book (Pringle, 2000), "If you were a bat you could hang by your thumbs and stay up all night," both of which strongly appeal to children. So it teaches them awareness of audience, that they need to appeal to the [reader].

3. Expanding Writing Project Expertise. Survey results indicate that participants draw from the AWC in various NWP leadership and outreach programs: invitational summer institutes (the center-piece program for preparing teacher-consultants), advanced institutes, and professional development programs for other teachers. Respondents reported using the AWC as tools for examining and evaluating student writing, selecting local anchor papers, and scoring student writing. A number of scorers incorporate the AWC into young writers' camps and parent and family workshops. Some savvy respondents use the AWC as a resource for marketing professional development in schools.

4. Teacher Ownership. The AWC embodies the authority of a rigorous assessment system, built with teacher input and with research as its primary focus, an area in which teachers are frequently disenfranchised. Within the first few hours of their initial training, teacher-scorers begin to wrestle with the concepts, making them their own, taking on the role of owner and co-author, doing the mental work to understand the system, questioning what they see as discrepancies, and creating niches for further inquiry into one or another facet of the system. Some tell us how the AWC operates as a silent partner as they write. One such teacher said:

> The [AWC] continuum helps me organize that complexity and chaos that is often writing. And I know it helps me as a writer. When I sit down to write something and I typically just write, write, write, just bluh on the page to begin with. But when I go in to look at it again, to do some revision, I don't pull the rubric out necessarily, pull that continuum out, but I often think about, okay, let's think, do each of my sentences flow nicely into one another? I really do. I really internalize I guess you could say those characteristics.

The scoring impact study investigated the ways in which individual scorers adopted the system. The following section sheds light on some organized efforts to use the AWC to meet local needs across the country.

A Common National Framework Goes Local

It is inherent in NWP culture that whatever is learned that is valuable goes home with teachers. This was true in the heyday of portfolio assessment, in the glory days of writing across the curriculum, and in other waves of promising instructional practice. It has been and is true of the AWC system. The dual nature of the system, first the standards that are shared across the

national network and second, the invitation to make use of the system in local contexts, creates a cyclic flow of adaptation from the national network to local sites and classrooms and back to the national network to begin the cycle again. More than two dozen writing project sites are at any one time actively pursuing locally based work using the AWC assessment system. Currently local uses of the AWC system range across four purposes illustrated here with examples.

Research

- At Texas State University–San Marcos, the writing project site conducted a 2-year qualitative study, *Assessment to Instruction*, investigating various classroom uses of the AWC. This study is presented in more detail below.
- Using returned state-scored assessment papers, researchers have conducted multiple studies of how the AWC system and various state rubrics are (or are not) aligned, many with the belief that the AWC may have implications for classroom use. A classroom teacher in the Upper Peninsula of Michigan has investigated the relationship between the AWC and the state assessment (including instructional usefulness of the information provided) as her master's thesis. This study by Sabin and Hetherington is also described below.
- A teacher-researcher from the University of North Texas is studying the AWC as a catalyst for the development of a common language about writing and for the improvement of student writing in his middle school classroom.
- Researchers from Mississippi State University and Auburn University have joined together to research prominent features of student writing and how those features correlate to the attributes of the AWC, with the goal of signaling readiness for instruction (Swain, Graves, & Morse, 2010).

Assessment

- Ball State University's writing project site is taking the AWC into the university setting, beginning with the selection of anchor papers for freshman English courses.
- Following Hurricane Katrina, when the state did not offer writing assessment, Mississippi sites, through their state network of writing projects, offered writing assessment for schools using the AWC as the analytic tool.

- The Mississippi Writing/Thinking Institute, through a subcontract with the state-selected assessment company, has designed a new state rubric based on the attributes of the AWC. Writing project teachers also select anchor papers and design and conduct training for scorers for the statewide writing assessment.

Professional Development

- The Hawai'i Writing Project is working with the Hawai'i State Department of Education to offer professional development connecting the Common Core Standards, the AWC, and writing prompts that will be offered to teachers online.
- An ongoing teacher inquiry group at Boise State University is exploring questions developed around the use of the AWC in classrooms. Teachers are writing a book to share their findings.
- At Canisius College in Buffalo, New York, a group of writing project teachers is investigating the AWC in a school partnership setting that serves refugee children from Africa.

Mentorship

- At the University of Southern Mississippi, the mentor for National Board for Professional Teaching Standards candidates uses the AWC to help candidates improve their writing, deconstructing it to better understand the scoring system and the underlying implications for quality writing.

Inside a local AWC project: Teacher inquiry at Texas State University–San Marcos. Lori Assaf, a leader in the Central Texas Writing Project, facilitated a 2-year AWC inquiry, Assessment to Instruction (A2I), with a core group of teachers. At the outset, teachers considered how they might make use of the AWC in their classrooms, answering questions such as, "How has the AWC influenced you and your understanding as a writing teacher?" "How are you navigating the aspects of good writing as listed on the AWC?" "What issues, constraints, problems are you experiencing related to writing in your classroom?" "How do these issues relate to the AWC and using it for instruction?" Unlike the National Scoring Conference where the primary goal is obtaining reliable scores, the goal of projects like this one is professional growth for teachers. When teachers gather to pursue their own learning, there is less emphasis on reliability and therefore less time spent calibrating and more time spent digging under the meaning of the terms and descriptors, more time for questioning, discussing, and making

applications to instruction. Early in their inquiry, these teachers found that using the AWC gave their students a common language for talking about writing and for thinking about how they might improve their writing. Each teacher worked independently to find the "best" way of using the AWC with her own students while, as a group, they shared and adapted ideas and constructed a set of principles for using the AWC for instruction. A handout used by the group in a conference presentation follows (Fig. 2.3).

1. Use LOTS of authentic dialogue. In order to make sense of the NWP-AWC and effective writing, students and teachers must talk with each other and other authors.
2. Teachers should model, demonstrate, and provide direct instruction for each NWP-AWC criteria and use high-quality literature to illustrate effective writing.
3. Students should identify each criterion by name and practice each criterion in their own writing.
4. NWP-AWC is a strength-based tool. It can help teachers identify where their students need to progress and it allows teachers to meet students where they are. It is the foundation of formative assessment as well as summative assessment.
5. The NWP-AWC should not be used as a rigid tool but as a flexible continuum. Make the NWP-AWC your own! Don't let it dictate your teaching or limit your students' writing.
6. Students should self-evaluate and self-assess in order to improve their writing.
7. Students and teachers should continuously look at their identity as writers and explore their strengths as developing writers.
8. Study and evaluate a piece of writing and not the writer.
9. Students and teachers need to be allowed to take risks.
10. Develop authentic writing and create opportunities for students to publish.

Fig. 2.3. Top Ten Principles for Using the Analytic Writing Continuum (as developed in the A2I project).

Northern Michigan University, Marquette writing project researchers compare the AWC to the MEAP. Inviting local teacher-consultants to be trained as scorers, Jan Sabin, director of the Upper Peninsula Writing Project at Northern Michigan University, investigated the extent to which the AWC maps onto the writing assessment within the Michigan Educational Assessment Program (MEAP). With training provided by NWP, Sabin's group used a sample of 225 state-scored papers written by elementary and middle school students for the comparison. Scorers calibrated to the AWC system and then scored the state papers. Analyses of the scores confirmed high correlations between five of the six attributes (Conventions being the exception) of the AWC and the MEAP holistic score (Hetherington, 2010), indicating that the details provided in the AWC descriptors could help teachers plan instruction that would help students to improve their writing and that would be responsive to the MEAP assessment. Based on these findings, Sabin's group of teacher-consultants is planning professional development to be offered to local schools.

Accessing the AWC. Is the AWC so powerful that one has but to hold it in hand to elicit influence on the teaching of writing? No, absolutely not, but the AWC does come with the caveat that to "hold it in hand" one should first come to understand its underlying tenets. Becoming familiar with the instrument and adopting an inquiry stance toward its applications are important first steps into the AWC community of use. The AWC policy suggests that entry into the community can be offered by the NWP itself or by any of the NWP sites having a cadre of teachers with AWC experience. NWP's network of sites across the country is the pathway for others into the community and to ensure its accessibility.

REFLECTIONS ON IMPACTS AND INFLUENCES

On reflection, one might ask, "Why does this particular system hold so much promise for assessment and the teaching of writing?" The answers are not totally elusive. First, the AWC system does what LeMahieu and Friedrich described in 2007: It "draws upon teacher research traditions to build an inquiry community that matures over time, in terms of both its substantive sophistication of and the inherent expectations employed in its assessments" (p. 102). Built within a culture of educators who value inquiry, the AWC continues to support both reflective practice and rigorous research.

The Scoring Impact Study confirms that those who participate in the national scoring conferences deepen their understandings of writing itself while reflecting on the particular qualities of the AWC. Some of these teach-

ers credit the AWC's value to its focus on "the writing" as opposed to "the writer." When a teacher focuses on "the writer" as she initially reviews a paper, the individual student enters into the scoring activity, sometimes resulting in lowered expectations ("but this is good work for him," or "it's better than her last piece") or other thoughts that diminish instruction. A focus on "the writing" rather than "the writer" provides clarity on the locus of authority. Teachers focus on what is on the page—the writing—and once that is understood in terms of the attributes and what would be required for improvement, the teacher, now with full awareness of the potential of the writing, then becomes the agent for bringing the writer, the student, back into the instructional process. Expectations remain high because the potential for the piece of writing to become better is not compromised. Others say that the focus on strengths, looking for "what is present" rather than "what is not there," enables a teacher to identify building blocks on which to layer instruction. They tell us that focusing on strengths by identifying where writing skill falls along the continuum makes it possible to identify next steps toward improving the writing.

The many local uses that have shown the flexibility of the AWC also point to the promise of the system. Researchers in New Paltz, New York, for example, translated the AWC into Spanish to better understand the writing of migrant children. Researchers at Clemson University have submitted elementary on-demand writing along with writing composed in naturally occurring classroom contexts. These researchers want a Vygotskian view of what students can do "in cooperation with others" within the classroom compared to how well they write in on-demand situations (Kaminski & Hunt-Barron, 2010). Researchers such as these, so far apart in miles, come together around the shared values of the AWC system and pursue their own interests while learning from each other.

Huot (2002) argues that assessments should be site-based, locally controlled, context-sensitive, rhetorically based, and accessible. The AWC system is *site based*, developed in response to a specific need of one organization that holds common values and beliefs toward writing. The system is *locally controlled* in that the NWP itself is responsible for revising, updating, validating, managing, and opening the system for review. It is *context sensitive* in that it honors the instructional goals and objectives of teachers, students, and schools across the country, while respecting local cultural and social environments. It is *rhetorically based* in that it adheres to recognized rhetorical principles and interpretation of texts. Finally it is accessible as shown by the AWC activities locally designed and purposed. Thus, studying the impact of the AWC system allows exploration of the possibilities for assessment that is locally contextualized yet linked to a common national framework and standards of performance.

ENDNOTE

1. Anne DiPardo, University of Colorado at Boulder; JoAnne Eresh, Achieve; Sandra Murphy, University of California, Davis; Gail Offen-Brown, University of California; Faye Peitzman, UCLA Graduate School of Education & Information Studies; Melanie Sperling, University of California, Riverside; Barbara Storms, California State University, East Bay, retired; Paul LeMahieu, Carnegie Foundation for the Advancement of Teaching; Sherry Swain, National Writing Project.
2. See also Chapter 3 of this volume (Persky).

REFERENCES

Culham, R. (2003). *Six + 1 traits of writing*. New York, NY: Scholastic Professional Books.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 64*(1), 5–30.

DiPardo, A., Storms, B. A., & Selland, M. (2011). Seeing voices: Assessing writerly stance in the NWP Analytic Writing Continuum. *Assessing Writing, 16*(3), 170-188.

Hetherington, J. M. (2010). *An analysis of the Michigan Educational Assessment Program rubric and the National Writing Project Analytical Writing Continuum* (unpublished master's thesis). Northern Michigan University, Marquette.

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

Kaminski, R., & Hunt-Barron. S. (2010). *Evaluating project WRITE: The teacher and student outcomes of a professional development program focusing on core components of writing workshop and the traits of writing.* Berkeley, CA: National Writing Project.

LeMahieu, P. G., & Friedrich, L. (2007). Looking at student work to develop evaluative frameworks: Why . . . and as important, how? In A. Davies & K. Busick (Eds.), *Classroom assessment: What's working in high schools, Book 1* (pp. 103–127). Courtenay, BC, Canada: Connections Publishing.

NWP Research and Evaluation Unit. (2010, July). *Writing project professional development continues to yield gains in student writing achievement* (Research Brief No. 2). http://www.nwp.org/cs/public/print/resource/3208

Pringle, L. P. (2000). *Bats!: Strange and wonderful.* Honesdale, PA: Boyds Mills Press.

Singer, N. R., & LeMahieu, P. G. (2012). The effect of scoring order on the independence of holistic and analytic scores. *Journal of Writing Assessment.* http://journalofwritingassessment.org/contact.php

Swain, S. S., Graves, R. L., & Morse, D. T. (2010). Prominent feature analysis: What it means for the classroom. *English Journal, 99*(4), 84–89.

Swain, S. S., LeMahieu, P., Sperling, M., Murphy, S., Fessehaie, S., & Smith, M. A. (2010, May). *Writing assessment and its impact on scorers.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

White, E. M. (1994). *Teaching and assessing writing* (2nd ed.). San Francisco, CA: Jossey-Bass.

White, E. M. (2007). *Assigning, responding, evaluating: A writing teacher's guide* (4th ed.). Boston, MA: Bedford/St. Martin's.