Seeing voices: Assessing writerly stance in the NWP Analytic Writing Continuum

Anne DiPardo a, *, Barbara A. Storms b,1, Makenzie Selland c,2

a School of Education – UCB 249, University of Colorado, Boulder, CO 80309-0249, United States
b Department of Educational Leadership, California State University, East Bay, 25800 Carlos Bee Blvd, AE 250, Hayward, CA 94542, United States
c School of Education, University of Colorado, Boulder, CO 80309-0249, United States

A R T I C L E   I N F O

Article history:
Received 28 July 2010
Received in revised form 10 December 2010
Accepted 16 January 2011
Available online 10 March 2011

Keywords:
Writing
Rubric
NWP
Voice
Stance
Evaluation

A B S T R A C T

This paper describes the process by which a rubric development team affiliated with the National Writing Project negotiated difficulties and dilemmas concerning an analytic scoring category initially termed Voice and later renamed Stance. Although these labels reference an aspect of student writing that many teachers value, the challenge of defining the construct for assessment purposes remains fraught with difficulty. This account of the category-and rubric-development process and related research includes a rationale for the decision to negotiate these challenges, as well as a description of how category-definition and score-point language were formulated and progressively refined across an initial and subsequent year of rubric use, evaluation, and revision.

© 2011 Elsevier Ltd. All rights reserved.

Writer’s “voice” is an at once vexing and enduring notion, both widely critiqued and persistently indispensable. Even amidst the current emphasis on five-paragraph essays and measurable skills, teaching resources emphasizing writerly voice rise to popularity year after year (e.g., Dean, 2000, 2006; Kittle, 2008; McQuade & Atwan, 2009; Romano, 2004). The notion lives wherever the craft of writing is prized; Google “voice in writing” to discover a cornucopia of strategies intended to help
student writers “sound” like “themselves” on the page, as well as resources for adults seeking to free that misty-but-vital something imprisoned over years of formal instruction (Goldberg, 1986; Lamott, 1995). This staying power of the notion is particularly curious in light of its many critics, including composition scholars who have proclaimed “voice” a remnant of the passé notion of the essential self (Bowden, 1999; Harris, 1987; Lensmire, 1998). Even scholars focusing on voice acknowledge the construct’s slipperiness; in introducing her edited book on the subject, Yancey (1994, p. vii) acknowledges that the more she studied the literature on voice, “the less certain I became, and the less I actually knew.” While “voice” may be conceptually elusive, many readers continue to find it unmistakable—in the cadence of a favorite author’s prose, in a student essay that engages and delights, in the textual poetry of a well-crafted speech.

When the National Writing Project (NWP) set about designing an assessment tool that would provide information to funding sources and particular sites concerning student achievement in writing, the questions of whether and how to evaluate “voice” immediately emerged as both important and perplexing. Originally developed in conjunction with a national research initiative encompassing six NWP sites around the U.S. (NWP, 2008a, 2008b), the Analytic Writing Continuum (hereafter, “rubric”) was designed to be sensitive to local goals and concerns as well as consistent with NWP beliefs and principles. As part of the project, NWP elected to assess student writing through an annual scoring conference at which teachers would be trained to evaluate writing gathered from across study and control sites. Endeavoring to formulate assessment strategies that would ultimately serve both accountability and instructional-improvement purposes? (LeMahieu & Reilly, 2004; Shepard, 2000), the development team was committed to creating a scoring rubric that would strongly reflect NWP teachers’ goals and priorities. With multiple sites desiring feedback on progress specific to particular aspects of student writing, an analytic rubric seemed a clear choice over holistic scoring alone; and when investigators from the initial group of research teams indicated the fostering of student “voice” as an important goal of their sites’ efforts, it seemed equally clear that this would be among the categories assessed.

Teachers affiliated with the NWP have long proclaimed the belief that good writing emerges from authors’ authentic interests, convictions, and personalities (NWP & Nagin, 2003). Founded in 1973, just a few years after the landmark Anglo-American Dartmouth Conference where the metaphor of “voice” was first introduced into educational discourse (Bowden, 1999), NWP grew up in the heyday of expressionism and commitment to writerly presence (Elbow, 1973, 1981; Graves, 1983; Macrorie, 1980). Many years later, with NWP’s reputation as a premier professional-development network long established (Lieberman, 2006), its teachers continue to regard “voice” as elemental to successful writing. The term bubbles up in popular books and articles penned by NWP teachers (e.g., Ballast, 2007; English, 2007; Talbert, 1998) and remains central to what teachers seek in forming writing communities in NWP summer institutes. To ignore “voice” in an NWP rubric would be an affront to the organization’s enduring sensibilities and its participants’ aspirations; but how to define what the construct is for this purpose, let alone measure it?

This article provides an account of how these questions were addressed in the development of an NWP analytic rubric. After a brief exploration of the term “voice” and its attendant assessment challenges, we present research documenting how teacher-scorers understood and attempted to use category rubrics over two consecutive years (2005–2006) of an annual NWP scoring conference. We then offer an account of the evolution of NWP’s decision to rename this category “Stance,” tracing ongoing efforts to formulate definitional and rubric language designed to guide scorers’ attention toward textual features rather than a more subjective sense of a writer’s identity and intentions.

1. The problem of “voice”

While the notion of “voice” remains pervasive in teachers’ judgments of writing quality as well as many state-level assessments (Zhao & Llosa, 2008), debates continue to swirl around its precise definition and feasibility as a measurable dimension (Hirvela & Belcher, 2001). Historians of the term

---

3 We use “category” throughout to refer to each aspect of writing measured by the AWC, as this is the term used by the NWP.
have traced its emergence in classical rhetoric and theater, associating “voice” with notions such as Aristotle’s *ethos* (Bowden, 1999; Cherry, 1988; Morner & Rausch, 1991; Yancey, 1994) and the Latin *persona*, literally the “person behind the mask” (Barnet, Berman, & Burto, 1971; Beckson & Ganz, 1989; Gibson, 1969; Morner & Rausch, 1991). Only with the mid-20th century emergence of the New Criticism did the term come to designate the more text-based concerns of how authors convey their attitudes toward a subject through fiction, figures of speech, cadence, sentence structure, and rhetorical devices (Abrams, 1957; Richards, 1929). By the latter decades of the century—amidst social unrest generally and the blossoming of the expressivist writing movement particularly—“voice” was everywhere, whether referencing the human presence of individual writers, stylistic attributes of texts, or the desire of historically disenfranchised groups to be heard in a still-larger sense (e.g., Guinier, 2003; Mairs, 1997; Meacham, 2003; Schiwy, 1996). Even as late-century critics argued that authorial “voice” is strategically constructed rather than discovered deep within (Barthes, 1974; Bowden, 1999; Coles, 1978; Harris, 1987), the term was enduringly employed in composition classrooms and texts to refer to a reader’s sense of an author’s essential human presence.

While systematic research on written “voice” is best characterized as scant and inconclusive, scholars have occasionally attempted to formulate more precise definitions of the term and of how this sense of authorial presence is co-constructed by writers and readers. In a provocative qualitative study exploring perceptions of prose “voice,” Palacas (1989) found that readers tended to associate an abundance of parentheticals (“I think,” “let me go back to something I noted earlier,” “to my mind, a trustworthy account,” p. 31) with a strong sense of writerly perspective. Over a decade later, Helms-Park and Stapleton (2003) designed and employed a *Voice Intensity Scale* directing readers’ attention to four main textual components said to comprise “voice”: “assertiveness,” “self-identification,” “reiteration of a central point,” and “authorial presence and autonomy of thought” (p. 245).

Defining the relevance of “voice” in L2 writers’ development has remained a point of debate in both the conceptual and empirical literatures (Hirvela & Belcher, 2001). Some have argued that prevalent conceptions of “voice” in the U.S. reflect American individualism and naïve notions of singular identities (Bowden, 1999), valuing emphatic, revelatory writing reflecting assertiveness and candor (Carson, 1992; Hirvela & Belcher, 2001; Ramanathan & Kaplan, 1996). Ramanathan and Atkinson (1999, p. 48) argue that many cultures take a rather different view of communicative acts, prizing “the subtle, interpretive, interdependent, nonassertive, and even nonverbal.” While a few L2 researchers argue that a culturally sensitive conception of “voice” can usefully inform writing instruction for diverse learners (Coady & Escamilla, 2005; Ivanic & Camps, 2001; Prior, 2001; Spalding, Wang, Lin, & Hu, 2009), others cast a more skeptical eye on the efficacy of attempts to foster an L2 writing “voice” (Atkinson, 2001; Stapleton, 2002).

The extent to which scorers’ assessments of “voice” are associated with judgments of overall writing quality remains a particular focus of debate with regards to both L1 and L2 writers. In the aforementioned study focusing on non-native speakers of English, Helms-Park and Stapleton (2003) found no significant correlation between judgments of overall writing quality (measured by Jacobs, Zinkgraf, Wormuth, Hatfield, and Highey’s *ESL Composition Profile* [1981]) and the presence of “voice” (measured by the authors’ own *Voice Intensity Scale*). More recently, Zhao and Llosa (2008) adopted similar methods in a study of L1 writers’ New York State Regents Examination high school graduation essays. In their analyses, Zhao and Llosa compared scores for each of the four components of “voice” specified on the *Voice Intensity Scale* (Helms-Park & Stapleton, 2003) with scores for overall quality. In marked contrast to Helms-Park and Stapleton’s findings, Zhao and Llosa documented strong correlations with overall writing quality for all four “voice” components, with high scores for “reiteration of a central point” emerging as particularly predictive of overall quality. Zhao and Llosa maintain that the correlation between “voice” intensity and overall writing quality may have been skewed by Helms-Park and Stapleton’s use of the *ESL Composition Profile*, a tool that does not specifically attend to issues related to “voice.” Maintaining that control of “voice” is central to success in both L1 and L2 students’ writing, Zhao and Llosa call for further research to address the many voids in our current understandings of this amorphous yet widely valued dimension of effective prose.

Variously defined, much debated, and decidedly under researched, the notion of “voice” presents a host of conceptual and psychometric challenges. While abidingly committed to honoring NWP teachers’ interest in assessing those aspects of student writing that they deem most significant, the
rubric-development team approached the challenge of assessing “voice” with caution. As we detail below, the process of defining and measuring “voice” for large-scale assessment purposes proved far from simple or smooth, eventually yielding a decision to rename the category altogether; but even as the rubric developers re-termed the category Voice (Stance), they struggled to formulate definitional and rubric language that would both capture the spirit of NWP teachers’ interest in “voice” while directing scorers’ attention to particular textual features in student writing.

2. Building the NWP Analytic Writing Continuum

2.1. The NWP rubric development team

In response to the challenge of formulating an evaluation tool to evaluate student work (early elementary through high school) from its U.S. research-network sites, NWP convened a team in 2005 that brought together expertise in research, composition, and the development and implementation of large-scale writing assessments. Through a combination of face-to-face meetings (one team meeting each winter and at the summer scoring conferences), on-line interactions, and phone conferences, the team collaborated closely in formulating and fine-tuning assessment strategies—working as a whole for initial rubric development, and later dividing into early/upper elementary, middle school, and high school groups for the purpose of testing of the rubric and identifying sample papers to be used in scorer training. At an initial scoring conference, some team members were the lead trainers in grade-level rooms while others collected data by observing processes, conducting interviews, or reviewing scoring patterns.

As members of the rubric-development team, the first two authors of this paper approached our work from different angles of interest, expertise, and experience. A long-time writing instructor, English educator, and literacy researcher, Anne tended to focus on the degree to which emerging descriptions of the categories and score points might align with teachers’ perceptions of what matters in student writing; she advocated throughout the rubric-development process for the inclusion of a voice-like category and actively participated in discussions of how to define and assess it. A veteran teacher with long experience in school leadership and professional development, Barbara also appreciated the allure of the notion of “voice”; however, her participation in the development of other large-scale writing assessments had left her considerably more skeptical concerning its feasibility as a measurable category. During the team’s rubric-development team’s ongoing discussions of how to define and assess “voice,” Barbara often sounded a note of caution.

2.2. An iterative development process

We report here on the initial two years of an iterative process of rubric development, testing, data-gathering, and refinement. As Fig. 1 shows, the process encompassed multiple recursive cycles of creating, refining, and testing an evolving rubric that would provide consistent criteria for scoring multiple types of student writing at various grade levels. Throughout this recursive design process, the team regarded its working patterns as closely akin to collaborative action research—that is, shared inquiry with the goal of improved and mutually agreed-upon strategies for practice (Bray, Lee, Smith, & Yorks, 2000).

2.2.1. Defining “Voice”

At an initial meeting in the spring of 2005, the team began the development process by reviewing a then-current version of the 6 + 1 Trait rubric (NWREL, 2005). Featuring Voice as one of its traits, the rubric was already in wide use in classrooms across the U.S., a fact that reinforced the team’s belief in the enduring relevance of the construct and provided a sense of how the term was commonly understood among teachers. In the language of the 2005 6 + 1 Trait rubric, Voice was defined in terms of writerly authenticity that makes a strong impression on the reader:

Voice is the personality of the writer coming through on the page. It is what gives the writing a sense of flavor, a uniqueness, and gives the reader the feeling that the writer is talking directly
to her. A strong sense of voice demands that the writer make a commitment to the writing and write honestly with conviction. In a paper with strong voice, the reader will get a sense that someone real is there on the page, whether the reader knows the writer or not. (NWREL, 2005)

Acknowledging the profound value of the 6 + 1 Trait rubric for classroom- and school-level use, the development team pondered the creation of an instrument that would carry a comparable level of face validity for teachers while also fitting the requirements of large-scale direct writing assessment; that is, the development team sought to direct scorers’ attention to features of written text rather than subjective conjectures concerning an author’s personality and intentions. Formulating a definition of Voice that adequately reflected this intention, however, proved difficult. In the end, the NWP team’s 2005 definition was somewhat more text-based than the 6 + 1 Trait formulation, yet retained traces of the common tendency to equate Voice with the degree to which a reader can sense a human presence:

The Voice category describes how effectively the writer communicates in a manner that is expressive and engaging, thereby revealing the writer’s stance toward the subject. Voice is evident when a writer shows a sense of his/her personality through the writing. (Analytic Writing Continuum, 2005)

Reluctant to stray too far from commonly accepted notions of “voice,” the development team decided to pilot an initial rubric based on this definition at the initial NWP scoring conference in 2005 and to make later adjustments based on scoring data and user feedback. As the team then turned to the task of developing score-point language, they engaged in an extensive process of reading student papers, searching for exemplars that might be considered emblematic of various levels of strength (e.g., “anchor papers”) to illustrate each score point during training.5

---

4 The 6 + 1 Trait definition of “Voice” has since been revised to read as follows: “Voice is the writer coming through the words, the sense that a real person is speaking to us and cares about the message. It is the heart and soul of the writing, the magic, the wit, the feeling, the life and breath. When the writer is engaged personally with the topic, he/she imparts a personal tone and flavor to the piece that is unmistakably his/hers alone. And it is that individual something – different from the mark of all other writers – that we call Voice” (NWREL, 2010a,b).

5 The other categories addressed (following the 6 + 1 model) were Content, Structure, Sentence Fluency, Diction, and Conventions.
Since a “5” score (on a 6 scale) reflected strong performance on a given category, the rubric language describing this score point provides a useful window onto how the Voice definition was taken up in the 2005 scoring guide (for the full text of the 2005 Voice rubric, see Appendix A):

A sincere attempt has been made to address the purpose and audience for the writing in an interesting way. It skips a beat here and there, however.

a) It’s a strong attempt although the best moments fade in and out
b) Moments of insight make this piece come alive.
c) The writer pays attention to which tone is best used on this piece. It’s not totally consistent but leans in the right direction.
d) Narrative writing has many moments when the writer feels connected.
e) Expository or persuasive writing leaves the reader with a sense of why the writer chose these ideas.
f) The voice is strong throughout the pieces, but the writer slack off a bit here and there.

While this rubric language hints at the team’s desire to direct attention to discursive features rather than authorial intention, consistent with prevalent conceptions of “voice,” references to “the writer” abound while readers’ impressions seem to supersede textual analysis. Though less than satisfied with this formulation, the team looked to the inaugural scoring conference for a better understanding of how teacher-scorers would understand and assess Voice.

2.2.2. Scorer training

At both the 2005 and 2006 NWP conferences, extensive scorer training was designed to ensure consistent understandings of each of the categories assessed as well as consistent uses of the category rubrics (later reflected in inter-rater reliability statistics). The training of scorers followed a trainer-of-trainers model commonly used in large-scale writing assessments, with the scoring organizers training a group of table leaders, who then assisted in training and monitoring groups of scorers assigned to particular tables (Weigle, 2002). Working in four grade-range groups (lower elementary, upper elementary, middle school, and high school) for a full day, members of the research team trained the table leaders to use the category rubrics. The table leaders together read exemplar papers, scored them using the analytic rubric, and engaged in extensive discussion concerning the ways in which the papers demonstrated the rubric language at various score points.

After the table leaders were trained, the full group of teacher-scorers received an initial introduction to the category definitions and rubrics and were then separated into rooms by grade-range groups and assigned to tables (each overseen by a previously trained table leader). A member of the NWP rubric-development team led training in each of the scoring rooms; room leaders guided the process of studying the scoring guides, examining anchor papers, and looking at papers at different performance levels. Participants practiced scoring sample essays and discussing until consistent use of the rubric criteria was established. Throughout the process, room leaders monitored room- and table-level discussion and the degree of agreement on scores. During the two days of each scoring, room-level recalibration training was conducted after each break. Over the two years, a total of nearly 14,000 essays were scored across the four grade-level rooms (see Table 1).

Each year, scorers evaluated student writing drawn from NWP research-network sites as well as non-NWP control classrooms. While scorers did not necessarily see the prompts to which the students had written (those were site- and sometimes even classroom-specific and not available to scorers), the variety of types of writing assessed included persuasive, narrative, expository, and character sketches.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of pieces of student writing</td>
<td>7505</td>
<td>6493</td>
</tr>
<tr>
<td>Number of table leaders</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Number of readers</td>
<td>61</td>
<td>60</td>
</tr>
</tbody>
</table>
Table 2
Voice/stance study data collection and analysis.

<table>
<thead>
<tr>
<th>Type of data collected</th>
<th>Amount or N</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>Scores</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All scores for 7505 papers</td>
<td>All scores for 6943 papers</td>
<td>Score agreement</td>
</tr>
<tr>
<td>23% double scored</td>
<td>15% double scorings</td>
<td>Inter-rater reliability by category</td>
</tr>
<tr>
<td>3rd score where first two were not adjacent</td>
<td>3rd score where first two were not adjacent</td>
<td></td>
</tr>
<tr>
<td>Field notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full group discussions about voice</td>
<td>Full group discussions about voice</td>
<td>Coding for patterns in language related to Voice/Stance</td>
</tr>
<tr>
<td>Table level discussions about voice (all levels)</td>
<td>Room- and table-level discussions about voice (middle school)</td>
<td></td>
</tr>
<tr>
<td>Interviews</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 middle school scorers</td>
<td>4 middle school scorers</td>
<td>Coding of transcripts for patterns in language related to Voice/Stance</td>
</tr>
<tr>
<td>3 high school scorers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3
Summary of quantitative data.

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of papers</td>
<td>7505</td>
<td>6943</td>
</tr>
<tr>
<td>Total number of scores (including first, second and third readings for all traits plus holistic scores)</td>
<td>54085</td>
<td>44771</td>
</tr>
<tr>
<td>Total number of Stance/Voice scores (including first, second and third)</td>
<td>7727</td>
<td>6396</td>
</tr>
<tr>
<td>Percent of second readings</td>
<td>23%</td>
<td>15%</td>
</tr>
<tr>
<td>Stance/Voice agreement level</td>
<td>90%</td>
<td>88%</td>
</tr>
</tbody>
</table>

3. Scoring study

As part of the rubric-refinement process, the development team gathered data at the 2005 and 2006 scoring conferences with the goal of addressing the following questions: 1) To what extent were scorers arriving at consistent Voice/Stance scores for particular papers?, 2) How were scorers understanding the Voice/Stance category?, and 3) What were scorers taking into account in arriving at their scoring judgments? As detailed below, data were both quantitative and qualitative (see Table 2).

3.1. Quantitative data

At both of the scoring events, each piece of student writing was scored at least once, with a pre-determined percentage of randomly selected papers read by a second scorer. At least 23% of all papers in 2005 and 15% of all papers in 2006 were double scored; if scores of double-read papers were not in agreement as determined by exact or adjacent scores, the piece of writing was read again by third reader (see Table 3). Additional second and third readings were conducted by table and room leaders as time permitted. Inter-rater reliability was determined by the proportion of double-scored traits that were in agreement; agreement between two raters was defined as a match in scores or being one score point apart on the six point scale, while disagreement was defined as more than one score point difference between the two raters’ scores.7

---

6 In discussing data gathered across the two years of the study, we refer to the category as Voice/Stance; in discussing only 2005 data, Voice; and in discussing only 2006 data, Voice (Stance).

7 The authors thank Sela Fessehaie, Data Manager/Analyst in the Research & Evaluation Unit of the National Writing Project.
3.2. Qualitative data

At the 2005 and 2006 scoring events, Barbara collected field notes from whole-group trainings and from grade-level-group discussions. Throughout both scoring conferences, she also collected room- and table-level observation notes: In 2005, she conducted observations in each of the four scoring rooms, while in 2006, her observations took place only in the middle-school room. Each day after the reading, the NWP research group met to review what they had observed, with members recording notes on these conversations. After each scoring conference, Barbara reviewed her field notes and the notes from the summary meetings to identify patterns of questions and language in relation to the Voice/Stance category.

At each of the readings, Barbara also conducted audio-recorded think-aloud protocol interviews (Flower & Hayes, 1981). In 2005, the research team wanted to understand more about how scorers scored Voice to determine if there were differences across grade levels. At the 2005 scoring conference, six scorers (three from the high-school and three from the middle-school room) were randomly selected and pulled out of the scoring situation for individual conversations in which Barbara probed their understandings of Voice and approaches to scoring this category (see Appendix B for interview guide). These scorers took to a separate room the batch of papers that they were currently scoring, along with the rubric and any other aids that they had used in determining scores (e.g., anchor papers). Interviewees were each asked to read three to five papers aloud to the researcher, to name what they were noticing about Voice in each paper, and to explain how they arrived at their scoring judgments.

Since the interviews from 2005 showed no particular differences in how scorers at different grade levels used the rubric’s Voice category to score, and because the agreement level for the Voice category at the 2005 scoring was lowest in the middle-school room, the research team decided to focus on only the middle-school readers in the continued 2006 study of how the Voice/Stance category was being evaluated. In 2006, four scorers from the middle-school room were selected to participate in think-aloud interviews (Appendix B). Scorers were given a copy of the rubric and were again allowed to bring any other aids (e.g., anchor papers) that they had used in determining scores. These scorers individually read the same set of seven pre-selected papers aloud and shared their perceptions related to Voice/Stance, including identifying evidence of the category in each paper and describing what they considered in making their Voice/Stance scoring judgments.

Soon after the scoring conferences, Barbara expanded her field notes, wrote reflective memos, and read and re-read transcripts of the think-aloud protocol interviews. She developed coding systems to identify patterns across scorer comments related to how scorers used the analytic rubric; how they defined, referred to, and described Voice/Stance in general; and how they reached their scoring judgments. When the rubric-development team reconvened some months later, Barbara presented the findings detailed in the following sections, thereby moving the group to ponder anew the challenge of defining the Voice/Stance category and of providing clear rubric language and training.

3.3. Assessing voice/stance on the page

While the overall agreement levels for the category of Stance/Voice were strong across both years (90% for 2005 and 88% for 2006), the interview and observational data from 2005 suggested that scorers brought notions to their assessments that did not reflect the language of the rubric. In particular, the 2005 interviews revealed that the scorers looked for the degree to which a writer conveyed audience awareness and engagement with a topic, expressed a sense of personality that drew the reader in, and used other elements of writing to establish a sense of authorial presence (see Table 4).

As the rubric-development team reconvened early in 2006, they discussed the study findings (and their own observations) that the Voice rubric language invited scorers to imagine the writer behind the text and to focus in reader-response fashion on their own subjective reactions. In an effort to challenge scorers’ conception of Voice as the sound of an authentic author speaking, the team elected to rename the category Voice (Stance) and to formulate a new definition:
The Voice (Stance) category describes how effectively the writing communicates a perspective through an appropriate level of formality, elements of style, and tone appropriate for the audience and purpose.

The rubric development team’s decision to use Stance as a synonym for Voice emerged through lengthy discussion of the attributes—including stylistic strategies and tone, how and if a sense of writerly perspective is clearly established, and so on—that are commonly regarded as contributing to a sense of authorial presence. A term employed by discourse analysts in examining oral language that suggest a speaker’s attitudes, positions, and perceptions (e.g., Jaffe, 2009; Tracy, 2002), Stance tends to emphasize aspects of language rather than directly assuming attributes of human intention or personality. In the rubric, references to “the writer” were removed so that the revised score-point language referenced only “the writing” (see Appendix C).

While renaming the category on the 2006 rubric Voice (Stance) was a relatively modest change, the team hoped that along with revised definitional and score-point language, 2006 scorers would focus more on textual features (see Appendix C for revised category definition and rubric). For example, the new rubric asked scorers to assess if a perspective was “evident and clear” rather than judge how and if the author communicated “expressively and engagingly,” as in the 2005 version. Scorers were asked to judge the level of sophistication in style and the degree to which that style communicated an attitude or stance toward the topic. Assessment of a text’s level of formality was added as a measure to help scorers assess the degree to which an essay’s tone is appropriate to its purpose and audience. Thus, the emphasis in the 2006 rubric was on evidence of perspective, tone, and style in the writing.

The training for the 2006 scoring was also revised to emphasize characteristics of writing rather than imagined writers. Room and table leaders were instructed to give repeated reminders to be “careful not to talk about the person,” to curb instances of “reader response,” and to focus on “the writing” rather than the “writer” when making assessments about the quality of Voice (Stance). Additionally, in discussing anchor papers, room leaders and table leaders highlighted textual evidence of commitment to a perspective; the presence of tone created through the consistent employment of stylistic techniques; appropriateness of the level of formality; and the level of sophistication created through the use of textual devices such as questions, quotes, metaphors, and similes.

While some issues persisted across the two years, 2006 interviewees generally found the revised rubric clearer and more useful (see Table 5). In both years, scorers reported looking for a strong sense

---

The revised NWP rubric’s emphasis on establishing and sustaining a consistent “perspective” as an indicator of Voice/Stance bears certain similarities to the dimension “reiteration of a central point” on the Voice Intensity Scale (Helms-Park & Stapleton, 2003), found by Zhao and Llosa (2008) to be particularly predictive of overall writing quality.

---

Table 4
Scoring voice, 2005.

<table>
<thead>
<tr>
<th>Voice definition</th>
<th>Implicit questions</th>
<th>Quotes from scorers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement with topic and audience</td>
<td>Is the writer aware of the audience and does he or she make that awareness clear? Does the writer seem to enjoy the topic and bring it to life?</td>
<td>“The writer shows an awareness of audience.” “Intelligent engagement with the topic came out.” “Having a gut level reaction of whether it’s authentic.”</td>
</tr>
<tr>
<td>Convey perspective and provoke response</td>
<td>Does the reader get a sense that a real person with a unique personality is writing authentically? Does the writing make the reader react?</td>
<td>“Voice is the essence of personality.” “If I react when I read it, it’s voice.”</td>
</tr>
<tr>
<td>Orchestration of stylistic elements</td>
<td>Does the writer use precise and natural words? Does the writer employ sentence length that is natural and appropriate for the subject? How can I distinguish “Voice” from particular stylistic features?</td>
<td>“Voice is sparkle and word choice.” “Voice has a lot to do with sentence structure and length, cadence.” “Individual details are Voice; hard to distinguish from the other categories.”</td>
</tr>
</tbody>
</table>
of writerly engagement or passion as they formulated Voice/Stance scores—but where 2005 scorers searched for an elusive sense of an author’s personality and intentions, 2006 scorers were guided by the revised rubric and training to look for ways that a strong sense of perspective was accomplished through control of style and tone. Interviewees in 2005 identified two concerns with the rubric that appeared to be largely rectified the following year: difficulty assessing “awareness of audience” and in distinguishing Voice from other rubric categories (particularly Diction and Sentence Fluency). In response to these concerns, the development team removed all references to audience awareness in the 2006 rubric and revised the category description and score-point language to distinguish the category more clearly (see Appendix C).

In the sections that follow, we discuss in greater detail the issues and themes that recurred across the two years of interview data with regards to scorers’ uses of the Voice/Stance rubrics.

### 3.4. Engaging with topic and audience

In both years of this study, scorers endeavored to assess the extent to which each essay conveyed a sense of commitment to communicating a given topic or position with rhetorical effectiveness. As 2005 scorers attempted to evaluate essays for evidence of writers’ “audience awareness,” they struggled to arrive at judgments of how well these anonymous writers were imagining their readers. With the 2006 rubric’s emphasis on evidence of how a given Voice (Stance) is established stylistically, interviewees’ descriptions of their scoring processes suggested closer attention to textual features.

Frequently referencing the rubric phrase “attitude toward the subject,” 2005 interviewees spoke of Voice in terms of the extent to which a writer communicates both awareness of audience and a sense of personal investment in a topic. Half of the six scorers interviewed mentioned a strong “awareness of audience” on the part of the writer, describing those instances when writers “talked” to the audience or “thought” of the audience as indicative of “Voice.” One scorer described this as a sense that the writer was “actually talking to me or someone else.” A scorer who gave a paper a “3” in Voice described the writing as “basic” because it had “a little attitude and awareness.” This assessment aligned with the 2005 rubric’s stipulation that a “3” in Voice may “show some awareness of the audience.”

Five of the 2005 scorers interviewed described looking for a writer’s embrace of the subject, noting that writers with Voice seemed to “buy in” to the topic, showed an “attitude toward the subject,” or
made “the topic their own.” One scorer described this quality as “not the content but how [the writer] talk[s] about it.” Although scorers were instructed not to consider writing prompts, another scorer lamented that some topics were not conducive to strong personal investment on the part of the writer because they were not “personal.” Acknowledging that “content makes a huge difference,” one scorer observed that even if a writer did not demonstrate a “personal” attitude toward the topic (an important exception concerning the presence of more explicitly academic prompts), the extent to which an “intelligent engagement with the topic [came] out” would affect a Voice score. Only two scorers also made Voice decisions based on how well the writer appeared to maintain a consistent attitude toward the subject.

While 2006 scorers were somewhat less concerned about the writer’s engagement with a topic, two of the four interviewees reported scoring somewhat higher when they sensed commitment or passion. Two noted that writers are more likely to register engagement with prompts that they find interesting; this could create a stronger sense of purpose, they suggested, thereby leading to the kind of increased precision in tone that creates Voice (Stance).

3.5. Conveying perspective and provoking response

Not surprisingly, the 2005 Voice category definition—emphasizing a writer’s ability to convey a sense of “personality” through “expressive and engaging writing”—elicited subjective speculations concerning authorial intention. As the category description was revised for the 2006 scoring to emphasize a sense of “perspective through an appropriate level of formality, elements of style, and tone” (an emphasis further underscored in table-leader and scorer training), interviewees grappled with the question of how a given perspective was accomplished stylistically. While 2005 interviewees looked first to their own emotional responses in detecting a human presence in the writing, 2006 interviewees endeavored to identify textual features that worked together to produce a sense of Voice (Stance).

The terms “authenticity” and “personality” recurred frequently across the 2005 interviews. Interviewees described looking to “learn about the writer” through the writing, characterizing Voice as a quality that let them “know students.” The six readers interviewed placed strong emphasis on locating a “unique person” who was writing from an “honest place,” evident in statements such as “Voice is the essence of personality.” Scorers described having “a gut-level reaction of whether it’s authentic,” of being able to “feel a precious soul,” and of being able to see “when it seems like [the writers are] behind their words.” One even described being more “generous” with texts when “I trust the writer,” while another noted, “If I giggle, then there’s Voice.” Interviewees sensed feelings such as “the writer’s excitement” and assessed Voice based on their ability to understand “the kind of person, attitude, [and] intellectual ability” that generated the words on the page. Conversely, interviewees also described times when they felt that “anyone could have written” a piece, that a writer was “faking it” or possibly “didn’t have time to be emotionally honest.”

Scorers interviewed in 2005 also defined and scored Voice through their assessment of a writer’s risk-taking. One scorer defined Voice as “a risk with who you are,” as “not worrying about what others want,” and as a space where a writer’s “own inner being comes through.” For this reader, a piece of writing that scored high for Voice was one in which the writer “[took] risks and let feelings and emotions show”; those receiving lower scores, on the other hand, couldn’t yet “trust their own voices.” Those tentative writers were often described as playing it safe, as refusing to “put [themselves] out there,” as “not committed,” or as willing to offer only “what children think adults want to know.” This notion of Voice as an ability to take risks sometimes coincided with a developmental perspective, as scorers looked for the potential for growth or for writers who showed unusual promise “for their age.” One scorer believed, for instance, that “when the writer takes ownership, then Voice emerges”; another, that “if it’s personal and unexpected for this age, then it’s Voice.”

Interviewees in 2005 also arrived at Voice scores by gauging their own levels of involvement. For example, one scorer noted that a writer “engaged me as a reader,” another that “If I react when I read it, it’s Voice, but if my expression doesn’t change, then the paper doesn’t have Voice.” Similarly, one interviewee knew that a piece of writing was “weak in Voice because I was bored; I had to force myself to finish.” Acknowledging the subjectivity of her judgments of Voice, yet another scorer observed that “I get a feeling from what [the writers] talk about. It lets me know students.” One scorer even created
a reader-response supplement to the rubric, deciding that a score of “5” meant “I’m hearing you,” a “4” that “I’m connecting with you,” and a “3” that the writer was “trying to do what the teacher asked them to do.” Taken together, these comments reflected a marked desire to feel affected by writing that conveyed a sense of personal conviction and presence.

The four scorers interviewed in 2006, on the other hand, employed the revised rubric language as they spoke of the quality of the writing’s perspective, style, formality and tone. Adopting a strategy introduced in the 2006 scorer training, readers used a technique of “naming” the perspective of the writer, such as “authoritative, humorous, sincere, or passionate.” As one scorer described it, this process of identifying the perspective helped the scorer to “detach from [an] emotional response” to the piece and to focus on text. Three analyzed perspective through the presence (or lack) of evidence to support a writer’s main point, while one looked for adequate “descriptors” to denote the presence of a consistent perspective and purpose. These four scorers also considered the degree to which a given perspective was sustained stylistically throughout a given essay.

However, even with their increased attention to the rubric language, the 2006 readers were not completely free of a tendency to rely on personal responses or to make judgments about a writer’s level of authenticity. Two of the readers interviewed looked for Voice (Stance) as the popularly conceived “human presence in the writing.” One reader recalled times when “I’m not relating to this author” versus sensing the presence of a “real” kid in the writing. This scorer noted that fostering Voice (Stance) was a process of “getting a kid to like themselves,” and therefore made better sense as a long-term learning goal rather than a one-time assessment focus. She explained that she would score a “5” for Voice (Stance) when she felt she was “sitting down and talking to the person,” whereas she would give a “1” or a “2” score when she felt that the student was attempting to “placate the teacher.” This scorer suggested that future versions of the rubric should include the phrase “human element,” an idea that aligned with her thoughts that “Voice is the person, the human element of the writing coming through, not just the mechanical element.” Another scorer, who described Voice (Stance) as the “oral component of writing,” noted that he could tell when a piece of writing was done “just to complete an assignment” and how Voice (Stance) came through when authors showed that they felt strongly about their topics.

3.6. Orchestrating stylistic elements

While interviewees in both years saw Voice/Stance as something greater than the sum of its parts, 2005 interviewees described particular challenges in conceiving of it as completely distinct from the other five scoring categories. Further, although scorers could make a large-grain-size distinction between the presence or absence of Voice/Stance, they often struggled to distinguish between adjacent score points. Although rubric and scorer-training changes addressed these issues to an extent, some degree of concern persisted.

3.6.1. Distinguishing from other scoring categories

Interviewees in 2005 reported particular difficulty in imagining Voice as somehow separate from the other categories. Content came into play in their scoring decisions, since “individual details are Voice” and there “less Voice in some writing because of the topic.” Formulaic Structure, as in the rigid five-paragraph essay “hamburger” approach, resulted in lower Voice scores, while multiple errors in conventions could make writing a voice-less “struggle to read.” Most often invoked in judging the presence of Voice were Diction and Sentence Fluency, as interviewees felt that they could “hear” a sense of Voice when there was a particularly effective use of vocabulary or an appealing flow to the sentences. One noted that Voice has a lot to do with sentence structure, length, [and] cadence,” and also that “Voice and word choice are closely associated.” Interviewees reported looking for “naturalness” in word choice, for precision in language, and for word choices that displayed emotion. Diction was also seen to affect the level of informality in the writing, which scorers felt “ad[ed] to Voice.”

In 2006, while readers did not specifically mention difficulty distinguishing the category, all spoke of attending to various features of text in arriving at Voice (Stance) scores. Scorer decisions were still affected by control of conventions, although most readers tried diligently not to consider grammar and punctuation errors when making Voice (Stance) judgments; one scorer, for instance, described
Voice (Stance) as “the [overall] what, not the [particular] how” of writing. Although three out of four scorers described errors in grammar and conventions as “a distraction,” they also felt that problems with conventions only served to lower a “Voice (Stance)” score when a reader couldn’t get through a paper.” One reader noted that a lower score for “Voice (Stance)” based on poor structure was “unfair” yet “unavoidable.” Although the 2006 scorers interviewed noticed singular elements of a piece of writing as they looked for evidence of Voice (Stance)—words, phrases, even punctuation that marked a piece as “unique”—they understood the category as the result of multiple elements working together. As they sought to determine the degree to which a writer was employing “sophisticated style to communicate stance or attitude,” 2006 scorers looked both for particular features of writing and how these worked together to construct a sense of perspective.

3.6.2. Distinguishing score points

Given scorers’ continuing difficulties in arriving at precise conceptions of Voice/Stance and in fully distinguishing it from the other scoring categories, making fine-grained score-point distinctions proved persistently challenging. While scorers had little difficulty in determining whether a given piece carried a stronger or weaker sense of Voice/Stance, deciding where to place it across the 6-point scoring scale was another matter. In response to the 2005 interviews, the development team threaded language throughout the 2006 rubric that referenced a descending order of degree or intensity; that is, a paper receiving a 6 score in voice “consistently” demonstrates a perspective, a 5 “usually,” a 4 “often,” a 3 “seldom,” a 2 “little or no extent,” while a 1 “lacks an apparent perspective” altogether. While the 2005 rubric had included some of these distinctions concerning intensity and consistency, they were flagged more methodically in the 2006 rubric and training.

Even so, 2006 scorers described persistent difficulties in determining the difference between score points, particularly at the lower end. Where markers of Voice (Stance) were clearly present, scorers could more readily determine their sufficiency; as one 2006 scorer observed, “a single “good phrase” was not enough to “bump up a score.” In evaluating weaker papers, however, scorers reported uncertainty in determining whether Voice (Stance) was seldom present, minimally present, or completely lacking. Similarly, although the 2006 interviewees were strongly committed to the importance of “consistency” in style and tone, this too occasioned difficulty in distinguishing among adjacent score points. While the rubric-development team sought to address this difficulty through close study of anchor and training papers, the lingering challenge of identifying fine-grained degrees of Voice (Stance) continued to make score-point distinctions challenging.

4. Discussion

4.1. Rubric development

As noted earlier, the rubric-development team held several inter-related goals in formulating the Voice/Stance scoring category. First and foremost, the team endeavored to formulate a clear and concise definition of Voice/Stance as a textual dimension that would hold face validity for teachers. Further, rubric language and training strategies were designed toward the goal of encouraging scorers to grasp definitional boundaries and internalize score-point language, thereby producing a high degree of inter-rater reliability. Finally, the team hoped that the participating teachers’ scoring-conference experiences in reflecting upon and evaluating Voice/Stance would provide insights that would enhance their efforts in their own classrooms.

If one were to judge the effectiveness of the rubric on the basis of inter-rater reliability statistics alone, the 90% scoring agreement for the Voice category in 2005 might be interpreted as ample evidence of success. Interview data gathered at the 2005 event, however, suggested that scorers’ high degree of agreement did not necessarily reflect a shared understanding of the category construct or the rubric’s score-point language. As interviewees revealed the reasoning behind their scoring judgments, it became evident that many were referencing popular conceptions of “voice” that had more to do with imagining the intentions of authors as people than with textual analysis. On the one hand, these approaches were not entirely inconsistent with the language of the 2005 rubric Voice-category
rubric, which emphasized the ability to write expressively and engagingly, to maintain a consistent attitude toward a subject, and to infuse a tone that added interest in a manner appropriate to the purpose and audience (see Appendix A). Although scorers often cited places in a text to illustrate the presence of writerly personality, they more often defined Voice in terms of their general sense that a writer was “authentic” and engaging as well as engaged.

Although inter-rater reliability for the category renamed Voice (Stance) dropped slightly at the 2006 scoring conference (scorers were 88% in agreement), the team’s observations, interviews, and informal interactions with scorers suggested that they were endeavoring to understand and score this category in terms of textual features to a much greater extent than in the previous scoring. Granted, some readers preferred the 2005 category definition and rubric, which had more closely reflected familiar conceptions of Voice; and as Voice remained in the foreground of the revised language, these conceptions did not go away entirely.

Overall, comments during the training also suggested that the 2006 rubric was intensifying scorers’ attention to textual features. Although uncertainties with the new Voice (Stance) rubric surfaced—most notably in decisions between score gradations of “2,” “3” and “4,” or between the descriptors “seldom,” “often,” or “usually”—these very difficulties would seem to suggest closer attention to textual features; however, scorers continued to report challenges in entirely eliminating notions of “human presence” from their considerations. Although 2006 rubric and training revisions fostered a shared understanding of Voice (Stance) as a stylistic accomplishment, the category remained the most problematic with regards to definitional and score-point distinctions of the six measured. Further, since the scoring conference organizers did not have access to these student writers, to information concerning their linguistic backgrounds, or to the prompts to which they were responding, much remains to be learned concerning other factors that may inform students’ writing and scorers’ judgments.

4.2. Voice on the page: measuring that ephemeral something

As the development team renamed the familiar-but-slippery Voice category Voice (Stance), revising and then fine-tuning its definition as well as the category-specific score point language, inter-rater reliabilities ranged from 88 to 90% over the first two years of the rubric’s use. Along with returning room leaders’ qualitative sense that scorers generally developed a clearer grasp of the category than in the first year, these developments strike us as noteworthy and good.

On the other hand, the matter remains far from definitively settled. Even in the years since, with this category now renamed simply Stance, scorers regularly press trainers for a more precise sense of how “tone” and “style” produce a sense of “perspective.” Distinguishing between adjacent score points continues to occasion concern—how to decide, for instance, whether tone and style “convincingly” versus “adequately” demonstrate a clear perspective? Perhaps more to the point, even given the current rubric’s emphasis on characteristics of “the writing,” NWP scorers remain the kind of caring and committed teachers who cannot help imagining anonymous writers as students in their classrooms. As they read through stacks of unknown students’ writing, they often remember ones they have known—their personalities and proclivities, their academic and linguistic backgrounds as well as their passions, aversions, and aspirations. Like good teachers everywhere, these scorers can never fully refrain from visualizing the writers behind the words, the emotional landscape of young lives, from imagining ways to connect, engage, and move forward. This, as we noted earlier, is a psychometric problem—but it is also precisely the turn of mind that marks these scorers as the kinds of teachers students and schools most need. We meet them at each year’s scoring conference, these people who give up beautiful June days to sit hour after hour internalizing a detailed rubric and then scoring enormous piles of student writing, moving through this arduous process with good spirits, keen attentiveness, and a level of commitment that marks the enterprise as profoundly important. They leave exhausted, but also full of excitement about what they have learned and how the experience will inform their

---

9 Scorers generally reported greater confidence in scoring the other categories; inter-rater reliabilities for the other five categories were largely in the mid-90 percentile range.
formative and summative assessments of their own students’ writing, how they will share their new insights in conversations with students and colleagues back at their sites and schools.

While the members of the rubric development team understood and embraced the NWP’s need for accountability to its funding sources and feedback for research-network sites, they were equally aware of their responsibility to the participating teachers who would use scoring-conference insights as the tool to enhance their efforts in the classroom (LeMahieu & Reilly, 2004; Shepard, 2000). In the end, the team elected to include a Voice/Stance category not because it can be rendered statistically measurable in some failsafe way, but because it mattered a great deal to the research-network sites looking to assess the effects of instruction on targeted aspects of student writing. In conferring with known young people about their writing, these teachers may well have lapsed into the familiar tropes—“the voice here is so you,” and so on—while large-scale assessment of anonymous writers clearly called for a different approach. But lurking insistently behind their work as scorers was a time-honored NWP tradition: that even our most public writing is infused with personal style and meaning (see Elbow, 1994). Even as we dismiss as quaintly naïve the notion that “good writing should have a voice, and the voice should be unmistakably your own” (Baker, 1962/1985, p. 6), most of us do strive for this effect, if not this very reality. It quite understandable that these teacher-scorers would want their students to develop this ability to sound like themselves on the page (Elbow, 1981; Eliot, 1943)—and, in valuing this aspect of writing, to want to see it somehow represented in indices of growth and achievement, too.

In the end, the results of our efforts to guide readers’ gazes to textual features were satisfactory but still open to question—intriguing, but hardly a solid basis for arguing that a particular teaching strategy led directly to student results in this sticky matter of effectively communicating “a perspective through an appropriate level of formality, elements of style, and tone appropriate for the audience and purpose” (Stance rubric category definition, 2006–present). In conversations at scoring tables and over lunch, participants at NWP scoring conferences continue to turn over the question of precisely what it is that leads them to the conclusion that a perspective has been orchestrated stylistically. It is the complex, nuanced sort of question that NWP teachers love, drawing at once on their profound interest in language, teaching, and of course, young people. As Elbow (2007) recently argued, while such stylistic analyses “may be more authoritative . . . they often imply a misleadingly technical or impersonal stance toward language (p. 178). In the generative spirit of “embracing contraries,” Elbow maintains, we need not only such close consideration of text, but also enduring attention to “how language issues from individual persons and physical bodies and how the same words differ, depending on who says them and how” (p. 175).

“Never trust the artist,” D.H. Lawrence famously warned; “trust the tale.” (1923/1965, p. 2). A tempting epigraph for this piece, but not quite apt, for no amount of encouragement to attend only to words on the page could ever erase NWP teachers’ abiding interest in the people who create them. Terming the category Stance invites a strategic stepping back that becomes helpful in assessing the broad sweep of how well NWP teachers are affecting large groups of anonymous students over time. While this is importantly useful information for funding sources and teachers alike, it hardly eclipses these scorers’ abiding interest in fostering students’ ability to be regarded as somehow “present on the page.” An organization that has long placed its trust in the creative intelligence of good teachers, NWP chose in the end to measure (however imperfectly) this ephemeral something—call it Voice, call it Stance—that remains a centerpiece of writing teachers’ work in classrooms.

Acknowledgement

The authors thank Paul LeMahieu, Sherry Swain, JoAnne Eresh, Sandra Murphy, Gail Offen-Brown, Faye Peitzman, and Melanie Sperling.

Appendix A. 2005 Voice category rubric

Definition: The Voice category describes how effectively the writer communicates in a manner that is expressive and engaging, thereby revealing the writer’s stance toward the subject. Voice is evident when a writer shows a sense of his/her personality through the writing.
Appendix B. Think-aloud protocol guides

2005 think-aloud directions: As you may know, NWP is interested in learning more about how teachers know and define voice in student writing. We are very interested in how you are scoring the aspect [category] of Voice at this scoring time. I’m going to ask you to think aloud with me as you read a small number (3–5) essays. Some of these essays will be new to you and others you may have read before. As you’re reading, please say aloud what you’re thinking and seeing about Voice in each essay and what you think about as you come to your judgment about a score for Voice. Do you have any questions for me?

Follow up prompts:

- What specific sentences, phrases or words demonstrated Voice to you?
- Do you think your understanding of what Voice is has changed as you’ve read more and more essays? In what ways has it changed? What words or phrases are you finding especially helpful in defining Voice for you at this time?

2006 think-aloud directions: We are interested in learning more about what “Voice (Stance)” looks like in student papers. We are very interested in what you notice when you are evaluating “Voice (Stance).” I’m going to ask you to think aloud with me as you read a small number of (7) essays. As you’re reading, please say aloud what you’re thinking and seeing about “Voice (Stance)” in each essay and what you think about as you come to your judgment about a score for “Voice (Stance).” Do you have any questions for me?

Follow up prompts:

- What specific sentences, phrases or words demonstrated “Voice (Stance)” to you?
• What does “Voice (Stance)” look like? Show me something in this essay that demonstrates “Voice (Stance).”
• Has your definition of “Voice (Stance)” changed since the beginning of this scoring, if so, how.

Appendix C. 2006 Voice (Stance) category rubric

Definition: The Voice (Stance) category describes how effectively the writing communicates a perspective through an appropriate level of formality, elements of style, and tone appropriate for the audience and purpose.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Consistently demonstrates a perspective that is evident and clear.</td>
<td>• Usually demonstrates a perspective that is evident and clear.</td>
<td>• Often demonstrates a perspective.</td>
<td>• Seldom demonstrates a perspective.</td>
<td>• Contains little or no evidence of a perspective.</td>
<td>• Lacks an apparent perspective.</td>
</tr>
<tr>
<td>• Consistently employs sophisticated style to communicate stance or attitude.</td>
<td>• Usually employs sophisticated style to communicate stance or attitude.</td>
<td>• Often employs style to communicate stance or attitude.</td>
<td>• Seldom employs style to communicate stance or attitude.</td>
<td>• Contains little style to communicate stance or attitude.</td>
<td>• Lacks style to communicate stance or attitude.</td>
</tr>
<tr>
<td>• Consistently exhibits formality and/or informality appropriate to the purpose and audience.</td>
<td>• Usually exhibits formality and/or informality appropriate to the purpose and audience.</td>
<td>• Often exhibits formality and/or informality appropriate to the purpose and audience.</td>
<td>• Seldom exhibits formality and/or informality.</td>
<td>• Exhibits formality and/or informality that is inappropriate to the purpose and audience.</td>
<td>• Exhibits formality and/or informality that is inappropriate to the purpose and audience.</td>
</tr>
<tr>
<td>• Demonstrates a tone that adds interest to the topic and is appropriate for the purpose and audience.</td>
<td>• Demonstrates a tone that is appropriate for the topic, audience, and purpose.</td>
<td>• Demonstrates a tone that is appropriate for the purpose and audience.</td>
<td>• Demonstrates a tone that adds interest to the topic and is appropriate for the purpose and audience.</td>
<td>• Shows little, if any, awareness of the audience or purpose.</td>
<td>• Shows little, if any, awareness of the audience or purpose.</td>
</tr>
</tbody>
</table>

References


**Anne DiPardo** is Professor of Literacy Studies at the University of Colorado at Boulder. She is a past editor of the NCTE empirical journal, *Research in the Teaching of English*, and presently chairs the NCTE Research Forum and serves as President-elect of the National Conference on Research in Language and Literacy.
Barbara A. Storms has been active in curriculum, instruction, and assessment reform as a teacher and administrator in urban schools, program administrator/researcher at Educational Testing Service, and a college professor. She helped design and conduct the 1998 NAEP Special Study for Classroom Writing and state and national performance assessment research and development projects.

Makenzie Selland is a doctoral student in literacy studies at the University of Colorado at Boulder. Prior to pursuing advanced study, she taught middle-school language arts in linguistically diverse, high-poverty schools. Her research focuses on innovative approaches to teacher induction and mentoring.