

Measuring Civic Writing: The Development and Validation of the Civically Engaged Writing Analysis Continuum

by Linda Friedrich, WestEd and Scott Strother, West Ed

Abstract: As youth increasingly access the public sphere and contribute to civic life through digital tools, scholars and educators are rethinking how civically engaged writing is taught, nurtured, and assessed. This article presents the conceptual underpinnings of the National Writing Project's Civically Engaged Writing Analysis Continuum (CEWAC), a new tool for assessing youth's civically engaged writing. It defines four attributes of civically engaged writing using qualitative analysis of expert interviews and literature: employs a public voice, advocates civic engagement or action, argues a position based on reasoning and evidence, and employs a structure to support a position. The article also presents reliability and validity evidence for CEWAC. The study finds that CEWAC has a moderate to high level of exact agreement and a high level of exact or adjacent agreement. Covariation analyses showed that, even with similar scoring patterns, CEWAC's attributes hold at least a moderate level of independence. This evidence, coupled with robust qualitative evidence around reliability and validity, establish CEWAC's strong technical properties. The findings suggest that CEWAC can be used both in research and in the classroom to make visible attributes of civically engaged writing often overlooked in traditional assessment frameworks.

Keywords: public writing, civic engagement, writing assessment, rubric, reliability

Opportunities to engage in participatory politics have expanded significantly in the digital age (Smith, 2013). Although youth navigate changing settings, audiences, and purposes for writing, they do not automatically cultivate the civic dispositions and academic skills necessary for thoughtfully engaging across a wide range of media and public contexts (Cohen, Kahne, Bowyer, Middaugh, & Rogowski, 2012; Purcell et al., 2012). Such skills and habits of mind must be explicitly taught if youth are to participate fully and productively in public life.

Participatory politics differ from traditional institutional politics in that they are peer-based, interactive, and not guided by deference to traditional elites and institutions, such as political parties or editorial boards (Jenkins, Purushotma, Weigel, Clinton, & Robinson, 2009). Rapid growth, spurred by the development of digital tools, requires significant shifts in civic education—and in how youth's civic learning is assessed (Kahne, Hodgin, & Eidman-Aadahl, 2016). Preparation for civic and political engagement, including investigation, dialogue, circulation, production, and mobilization, must be taught differently because these skills are practiced differently (Kahne, Middaugh, & Allen, 2014). Guided by this conceptualization of civic engagement, the National Writing Project (NWP) and its partners have developed the Civically Engaged Writing Analysis Continuum (CEWAC), an analytic writing rubric that assesses youth's ability, both in academic and extracurricular settings, to engage in civic arguments about issues that are meaningful to them and their communities. This article defines four attributes of civically engaged argument writing based on this research. It also reports results from a validation and reliability study that surfaced the inherent tension between scoring youth civic writing within a scoring system and reading it within the context of one's own civic commitments and values.

Literature

Rationale for Assessing Writing as a Measure of Civic Engagement

Increasingly, youth engage in online political dialogue and action. Notably, the proportion of youth posting comments about political issues on websites or blogs grew to just under 30% in 2012 while 54% of those who use the Internet have engaged in online dialogue related to politics (Smith, 2013). As they interact, youth encounter numerous challenges—assessment of sources' trustworthiness, political conflict, uncivil/unproductive dialogue (Kushin & Kitchener, 2009), and racist statements and interactions that can have a profound impact on their civic engagement (Weinstein, Rundle, & James, 2015).

Youth must learn essential skills for engaging in public action both through online encounters and in classrooms. These settings thus become mutually supportive as civic engagement skills are learned and applied. Further, the Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and the College, Career, and Civic Life (C3) Framework (National Council for the Social Studies [NCSS], 2013) emphasize the centrality of young people's ability to engage in argument rooted in thoughtful consideration of evidence to their preparedness for college, career, and citizenship. In light of new standards and youth's growing civic engagement through writing, we argue that assessing youth's ability to engage in argument through civically engaged writing is direct and authentic and is likely to reveal the quality of the skills they bring to other civic processes. By civically engaged writing, we mean any public writing for an audience beyond the writer's immediate family and friends that focuses on civic issues of significance to the writer, the community, or the public.

CEWAC's Intended Use and Interpretation

With its longtime interest in the civic dimensions of writing, NWP and its partners designed CEWAC to analyze the quality of youth writing written for a public audience focused on civic issues that matter to youth, their communities, and the public. Civically engaged writing may include public letters, opinion editorials, petitions, videos, extended online commentary, and the like. CEWAC expands writing assessment beyond traditional academic essays and on-demand prompts to focus on publishable writing for an authentic public audience.

CEWAC is intended to be used for two purposes. First, it is designed for use in evaluations of programs that build youths' capacity for civic engagement through writing. These include both programs that directly serve youth (e.g., youth writing camps) and those focusing on the adults who support youth learning (e.g., teacher professional development). Specifically, CEWAC supports inferences about such programs' impacts on the quality of youths' public writing. CEWAC's use for program evaluation is the focus of the validation efforts reported here.

Second, CEWAC's language and supporting materials are designed to support assessment for learning with educators and youth (Stiggins, 2005). Its language can assist teachers in making curricular and instructional decisions (e.g., What instruction might support students in transforming academic research about civic issues into writing for a public audience?). Scoring drafts of students' public writing may surface areas where youth would benefit from additional instruction. CEWAC's language and annotated

exemplars have the potential to build youths' independence in analyzing the quality of their own work. The study reported here does not consider reliability or validity in these learning contexts.

Rationale for Analytic Scoring

CEWAC's potential classroom use motivates the choice to create an analytic, rather than holistic, scoring guide. Analytic scoring focuses on discrete dimensions of a written product (Finson & Ormsbee, 1998; Rezaei & Lovorn, 2010). Analyzing attributes of writing separately may allow for more comprehensive construct coverage, thereby increasing the validity of analytic scoring (Quinlan, Higgins, & Wolff, 2009) and have greater instructional benefits because it builds understanding of how writing can improve (Swain & Friedrich, 2012). An analytic rubric allows the measurement of both those writing skills that enjoy broad consensus as related to civic engagement (e.g., develop arguments based on reasoning and evidence) and those that are emergent (e.g., advocate for a specific civic action).

CEWAC's Development Process

Teaching civics differently in response to the changing landscape also requires new ways to assess students' learning. Because civically engaged writing is central to this new conception, NWP facilitated an iterative design process in collaboration with a 12-member working group that included scholars of writing and civic education, a psychometrician, and educators working in schools, districts, and education nonprofits. The working group guided CEWAC's content development. This research was determined to be exempt; all interviews were conducted with the informed consent of interviewees, and their identities were masked to preserve confidentiality.

Phase One: Construct Development

During the first development phase, we conducted a literature review and interviews (Berry, 2002; Mishler, 1986) with a purposive sample of 16 civic education experts with more than a decade of experience in research or teaching in civic writing (Patton, 2002). Interviewees were recommended by the advisory group and came from fields of composition and rhetoric, civic and social studies education, and youth development. We employed a semistructured interview protocol (Seidman, 2005) that invited interviewees to describe the qualities they hoped to observe in youths' civically engaged arguments and respond to an initial set of attributes. All interviews were audio recorded, transcribed, and coded for our initial constructs (Miles & Huberman, 1994). We then prepared data summaries that supported the revision of the conceptual framework, which was initially developed by the study's principal investigators and based on an initial literature review and the subsequent development of an initial rubric containing five attributes.

Phase Two: Rubric Revision

The second phase of development engaged the working group in applying the initial rubric, based on the revised conceptual framework, to a range of high school youths' writing about civic issues, such as letters to the editor, petitions, comments about civic issues on a public website, issue analysis papers, and on-demand source-based arguments. Our analysis revealed that CEWAC's five initial attributes exhibited significant overlap; therefore, we condensed and restructured them into a new rubric with a four-point scale. The first two phases ensured that CEWAC's constructs were appropriate to the content area and offered adequate construct coverage (Messick, 1994).

Phase Three: Anchor Paper Selection and Rubric Finalization

Phase three focused on developing materials for reader training. The working group selected anchor and calibration papers (Moskal & Leydens, 2000; Wiggins, 1994) for high-school youths' public civically engaged writing (petitions, public letters, and journalistic writing). We held an initial 3.5-day meeting to select and begin annotating anchors, during which we conducted three rounds of winnowing and sorting the different types of writing. In round one, subgroups did a rough sort of approximately 150 samples for one writing type. They set aside papers for which the subgroup could not come to consensus on whether it was high, medium, or low. In round two, a different subgroup provided preliminary scores for one type of paper, again setting aside papers for which the subgroup couldn't come to consensus. In round three, all working group members independently scored the remaining samples. We then discussed the working group's level of agreement about scores and the implications for anchor paper selection. In addition, we crafted annotations to explain the rationale for each assigned score. While selecting anchor papers, the working group further clarified attribute definitions and scaling language. (See Appendix A for CEWAC rubric and www.cewac.nwp.org for annotated anchor papers.)

Phase Four: Pilot Scoring

A pilot scoring session composed the fourth phase. Twenty readers, four table leaders, and two room leaders from 10 states participated. All readers were high school English language arts or social studies teachers with at least three years of teaching experience and were nominated by a local Writing Project site director, EL Education, or a member of the working group because of their experience with teaching civic writing. Readers engaged in nine hours of training and nine hours of scoring. They scored 553 papers composed by high school youth. Room leaders first guided readers in studying CEWAC's rubric language. Room leaders then introduced the anchor papers and explained the rationale for each of the four score points. Readers completed calibration exercises and then independently rated papers. To calculate the reliability of the scoring system, every paper was independently scored by two readers assigned to different table groups. Every score off by one or more points was independently scored a third time by an adjudicator.

In addition to tracking interrater agreement, the research team conducted think-aloud interviews with a stratified random sample of eight readers (DiPardo, Storms, & Selland, 2011; Wolfe, 2005). They were instructed to verbalize their thinking as they read aloud and scored one of two researcher-selected writing samples. Readers were also asked for feedback on CEWAC's language and training procedures. All interviews were audio-recorded and transcribed. We analyzed transcripts for use of rubric language and other training materials (i.e., anchor papers), non-evaluative comments, references to construct-irrelevant features of writing (e.g., grammar; Messick, 1994), and rationale given for scores (Wolfe, 2005). All readers completed a closing survey focused on CEWAC's perceived usefulness, independence of attributes, points of confusion or disagreement, and ideas for improvement.

Phase Five: Analysis of Scoring Data and Final Revision

Following the scoring conference, the research team analyzed the qualitative and quantitative data collected. We aimed to demonstrate the validity and reliability of the rubric and to identify areas for improvement. The findings, reported below, guided a final revision of the attributes on the rubric (see Appendix B).

For reliability analysis, we first computed Cohen's Kappa coefficients to analyze interrater agreement across all score points and for individual attributes. We then performed descriptive analyses to find the percent exact and adjacent agreement of the readers. We conducted chi-square analyses to analyze differences in agreement patterns across attributes. We employed descriptive and chisquare analyses to unpack the prevalence of adjacent agreement at each score point (1–2, 2–3, and 3–4) and to analyze how readers' scores deviated from the scores of the adjudicators.

To explore the independence of attributes, as well as uncover any issues with dependence, we performed covariance analyses. We subsequently reviewed the corresponding qualitative data where readers or experts discussed the independence of attributes. Validity was analyzed using the think-aloud interviews and closing survey. Additionally, we compared CEWAC to an external measure, NWP's Analytic Writing Continuum for Source-Based Argument (AWC–SBA), using a correlation matrix to explore the alignment of attributes.

Through this process, we developed the conceptual framework that underpins CEWAC and created and tested the reliability and validity of the CEWAC rubric.

Conceptual Framework for Defining CEWAC Attributes

CEWAC responds to the challenge of teasing apart the relationship between arguments developed for authentic public purposes and arguments written to demonstrate mastery of academic skills. To this end, CEWAC defines four attributes for assessing youths' civically engaged writing:

- employs a public voice (EPV),
- advocates civic engagement or action (ACE),
- argues a position based on reasoning and evidence (AP), and
- employs a structure to support a position (ES).

Employs a Public Voice (EPV)

Developing an effective public voice is critical to youths' engagement in civic life given how digital tools have expanded access to the public sphere (Levine, 2008; Rheingold, 2008). Levine (2008) argues that a public voice is "any voice or style that has a chance of persuading other people (outside one's intimate circle) about shared matters, issues, or problems" (p. 121). Building on Levine's notion, CEWAC defines public voice as being "directed beyond [the author's] immediate family and friends." Drawing from rich traditions in literacy studies and civic education, CEWAC conceptualizes public voice as an attribute of writing that youth can learn and make conscious choices about (cf. Fletcher, 2015; Levine, 2008; Sperling & Appleman, 2011). Like scholars of youths' use of new media, CEWAC's definition of public voice recognizes that writers assume different voices on different platforms and for different purposes and audiences (boyd, 2014; Gold, Garcia, & Knutson, 2019). As teachers work with youth in developing public, civic arguments, the EPV attribute's explicit naming of an "intended audience" can open up classroom dialogue about unanticipated and even hostile audiences (Gold et al., 2019).

CEWAC decomposes the EPV attribute into two threads: (a) employs rhetorical strategies, tone, and style to contribute to civic discourse or influence action, and (b) establishes a writer's credibility. The EPV attribute's first thread examines the appropriateness of rhetorical strategy, tone, and style for intended audiences and purposes (Lazere, 2005; Shresthova, 2013). When the purpose focuses on contributing to civic discourse, CEWAC emphasizes the use of tone, style, and rhetoric that create

openness and demonstrate respect when considering alternative viewpoints (Harris, 2006; Hess, 2009; Kahne & Middaugh, 2008; Youniss, 2012). At the same time, CEWAC “allow[s] for the centrality of emotionality, attachment, and relationship to move people to civic action” (CEWAC Working Group, 2016). The potential for emotional engagement, well established in rhetorical tradition, underscores one way in which CEWAC speaks back to practices in secondary writing instruction that advocate for particular formulas or prohibitions of the personal in academic argument.

In the crowded public (and often digital) square, writers compete for initial attention, but they also need to establish credibility. CEWAC emphasizes the importance of establishing credibility as a feature of a public voice. As one interviewee noted, “Sometimes there’s a compelling nature to the voice or tone or stance that says not only ‘I’m credible,’ but also ‘I’m worthy of listening to in a crowded marketplace of people pitching ideas and positions.’” CEWAC defines the establishment of credibility as gaining the trust and confidence of the audience through thoughtful choices about language and content, including personal stories, and the use of credible information and data. Including personal narratives is particularly important for youth in marginalized communities. One interviewee explained how his work “tries to highlight a particular point of view that might not be typically represented.”

Advocates Civic Engagement or Action (ACE)

Developing as an active and engaged citizen involves building a sense of efficacy and the skills necessary to take informed action and mobilize (Kahne et al., 2014). This attribute is aligned with the “skills and dispositions necessary for an active civic life” (NCSS, 2013, p. 59). CEWAC focuses on two purposes for civically engaged writing: raising awareness of issues and advocating action to address important public questions.[1] The most effective civically engaged writing can demonstrate both why grappling with a specific issue is of public importance and that the proposed solution represents the best course of action. However, some genres of civically engaged writing prize brevity (e.g., letters to the editor, petitions) and may give more weight to either awareness or action.

In order to raise awareness, youth need to analyze and develop a deep understanding of the issue and the problem being addressed (Rubin, 2012; Terriquez, 2015). The first thread focuses on how well a piece of writing articulates an issue’s civic significance and raises the public’s awareness of it. It evaluates how contextual information connected to the civic issue—explanations of origin, impact, and why it needs to be addressed—helps raise public knowledge about and awareness of the issue, and establish its significance (Kahne & Middaugh, 2008; NCSS, 2013).

For writing that advocates specific civic actions, CEWAC analyzes the reasonableness and feasibility of the proposed action (Lynch, George, & Cooper, 1997). Pieces that step beyond raising awareness “build bridges from voice to influence” (Kahne et al., 2016, p. 24). This may include analyzing the opportunities and strategies currently available to influence public policy (Ito et al., 2013; Soep, 2014). Some civic engagement scholars argue that engaging students in proposing action teaches youth that “change [is] possible” (Terriquez, 2015, p. 235). Two expert informants, however, cautioned *against* always requiring students to advocate for civic action, especially in schools where requiring action may breed cynicism rather than authentic engagement (Rubin, 2012). One reflected, “When kids are dealing with a problem that’s embedded in deep-seated structural inequality, I’d rather have them understand that than . . . [say] that they have to come up with a solution.” This caveat, coupled with a desire to assess a range of civically engaged writing, means that CEWAC is designed to analyze writing that contributes to understanding problems, as well as that which advocates for action.

Argues a Position Based on Reasoning and Evidence (AP)

While the first two CEWAC attributes emphasize the public and civic nature of writing, the third attribute—argues a position based on reasoning and evidence—bridges public writing and valued academic skills. Like academic argument rubrics, this attribute analyzes the quality of reasoning used to connect claims and evidence (Hillocks, 2011; Lazere, 2005). It also adds two civic dimensions related to reasoning and use of evidence: how value structures inform reasoning and how personal experience can function as evidence. In contrast to academic argument rubrics, which may require explicit treatment of counter arguments, CEWAC makes this thread optional given the brevity of much civically engaged writing.

CEWAC adds an explicit civic dimension to its analysis of reasoning by considering the value structure that guides the reasoning presented. As one interviewee commented, “Civic writing and other kinds of communication in the civic domain needs to be informed by facts and be responsible to the facts, but it’s [also] going to be about values.” All civically engaged writing, whether effectively or weakly developed, is framed by values and morals gained through upbringing, education, or lived experience. Indeed, civic positions originate in individuals’ and groups’ “real histories” (Lynch et al., 1997, p. 68).

A central activity in civic and political life is for citizens to hear multiple perspectives and reflect on varied viewpoints as they discuss issues of public concern (Kahne, Lee, & Feezell, 2012; Rheingold, 2012). Developing an in-depth understanding of civic and political issues requires seeking out and considering multiple perspectives informed by race, class, gender, age, ideology, and geographic location. Indeed, today’s youth are likely to encounter a range of communities with different values, beliefs, ways of thinking, and speaking (McWilliams, 2013). When writing includes alternative or opposing viewpoints, CEWAC advocates grappling with the complexity and nuance offered by differing perspectives rather than outright dismissal (Hess, 2009; Lazere, 2005; Parker, 2011).

Employs a Structure to Support a Position (ES)

When effective, the structure of an argument enhances the writing’s central message. This is true for public arguments, as well as for academic writing. Therefore, CEWAC’s final attribute—employs a structure to support a position—analyzes how organization and structure help develop the central argument (Culham, 2003). CEWAC’s approach to measuring structure adapts language from the NWP’s Analytic Writing Continuum (Bang, 2013; Swain & LeMahieu, 2012) to the context of civically-engaged writing. Thus, CEWAC focuses on the overall organization of the writing, how the opening and closing may enhance civic engagement, and the linkages among ideas.

Validity and Reliability

To explore the validity of CEWAC’s attributes and the reliability of its application to youths’ public civically engaged writing, we analyzed the quantitative and qualitative data collected during the pilot scoring conference. To demonstrate CEWAC’s reliability, we present findings about interrater agreement and adjudication. To assess validity, we consider the independence of attributes, evidence from the closing survey and think-aloud interviews, and comparison with an established measure of academic argument writing.

Interrater Agreement

Following the scoring session, reliability was first established through measuring interrater agreement. There was an exact or adjacent agreement rate of 91.9% and an exact agreement rate of 46.4%, which yields a Cohen’s kappa coefficient of 0.29. This agreement level was markedly consistent across all attributes (see Table 1), with Kappa coefficients ranging from 0.26 to 0.30.

Table 1 includes percent exact agreement, percent adjacent agreement (e.g., one reader scores an attribute as a 2, and the other scores it as a 3), and percent other (two or three apart, e.g., one reader scores an attribute as a 2, and the other scores it as a 4). The agreement of EPV and ACE scores trends slightly lower; however, there was not a significant difference in agreement patterns across attributes ($\chi^2 = 4.01, p > .05$).

Table 1

Agreement (%) for Double-Scored Papers, in Total and by Attribute

<u>Agreement</u>	<u>Total</u>	<u>EPV</u>	<u>ACE</u>	<u>AP</u>	<u>ES</u>
% Exact Agreement	46.4	44.3	46.5	47.2	47.6
% Adjacent	45.5	47.9	45.9	43.9	44.1
% Other	8.1	7.8	7.8	8.9	8.3

Our largest concern was the high percent of adjacent scores. We wanted to unpack where adjacent scores were happening and why. Readers had the most difficult time teasing apart the difference between 3s and 4s. Looking across instances where both readers scored an attribute a 3 or 4, they disagreed 40% of the time. This is higher than expected but significantly better than chance, $\chi^2 = 16.27, p < .01$, where the expected proportion of disagreement from random scores (with the same proportion of 3s and 4s) would be 47.2%. The same trend existed when looking across all instances where both readers scored an attribute a 2 or 3, where they disagreed 38.3% of the time (versus a chance value of 49.1%). There was slightly less disagreement between 1 and 2, where readers disagreed 30.6% of the time (versus a chance value of 44.4%).

Adjudication

To further understand interrater agreement, we analyzed how readers’ scores compared to the adjudicators’ scores (see Table 2). We wanted to understand whether readers were consistently scoring above or below adjudication. Adjudicators agreed with one reader on 76.2% of disagreements. The scores misaligned with adjudicators’ scores were most often one below the adjudicated score. This indicates that readers sometimes had higher criteria, thus trended toward lower scores than adjudicators. The pattern is consistent across attributes with no significant difference ($\chi^2 = 9.67, p > .05$).

Table 2

Pattern of Misaligned Readers' Scores Compared to the Adjudicator's Score

<u>Attribute</u>	<u>3 Below (%)</u>	<u>2 Below (%)</u>	<u>1 Below (%)</u>	<u>1 Above (%)</u>	<u>2 Above (%)</u>	<u>3 Above (%)</u>
Total	0.8	11.6	46.1	33.5	7.4	0.5
EPV	1.0	13.0	49.2	28.7	7.2	1.0
ACE	1.2	10.2	46.1	36.0	6.2	0.3
AP	0.8	12.0	44.0	35.1	7.8	0.3
ES	0.3	11.1	45.1	34.8	8.2	0.5

Independence of Attributes

The similar agreement and disagreement patterns across all four attributes indicate a potential lack of independence across attributes. This could stem from the attributes themselves; that is, they do not sufficiently isolate the concepts that they are attempting to measure. In addition, we recognize that readers' holistic perception of writing quality likely shapes how they score each attribute, rather than readers providing a truly independent score for each attribute. For us, the primary purpose of analytic scoring is to inform instruction and feedback.

An analysis of covariance explored this and affirmed some relationship between the scores, but it also indicates that there is meaningful independent variation for each attribute (see Table 3). The values were low enough to show a good amount of independence, considering some covariation is expected as each scale is only four points and writing samples typically have an underlying holistic quality that will cluster the attributes' scores more than chance alone.

Table 3

Inter-Item Covariance Matrix

<u>Attribute</u>	<u>EPV</u>	<u>ACE</u>	<u>AP</u>	<u>ES</u>
EPV	0.824	0.672	0.623	0.664
ACE	0.672	0.860	0.651	0.695
AP	0.623	0.651	0.798	0.663
ES	0.664	0.695	0.663	0.841

Validity Evidence from Think-Aloud Interviews and Reader Survey

Reader surveys and think-aloud interviews confirm the quantitative findings about reliability. Readers recognized some independence across the attributes but shared concerns around conceptual overlap. In response to a Likert-scale survey item, slightly more than half of readers reported each attribute as

independent from the others. The think-aloud interviews surfaced the complexities of scoring public, civically engaged writing. To illustrate this complexity, we focused on readers’ scores and rationale for scoring each of the four attributes (see Tables 4 and 5). In three think-aloud interviews, readers scored a public letter about “Renewable Energy.” The remaining five think-aloud interviews focused on the public letter, “Brutality Towards Police.”

Table 4

Think-Aloud Readers’ Scores on Public Letters—Renewable Energy

<u>Interview Number</u>	<u>EPV</u>	<u>ACE</u>	<u>AP</u>	<u>ES</u>
Interview 1	3	4	3	4
Interview 2	3	4	3	4
Interview 3	3	Missing	2	3

Table 5

Think-Aloud Readers’ Scores on Public Letters—Brutality Towards Police

<u>Interview Number</u>	<u>EPV</u>	<u>ACE</u>	<u>AP</u>	<u>ES</u>
Interview 4	4	4	3	4
Interview 5	4	4	3	4
Interview 6	4	3	2	3
Interview 7	3	4	2	4
Interview 8	2	3	2	3

EPV. In the closing survey, four readers (17.4%) indicated this attribute needed additional clarification. During the think-aloud interviews, readers named similar rationale and evidence from the writing for their EPV scores. However, the interviews reveal two complexities in scoring this attribute. First, readers viewed the attribute’s credibility thread as being closely related to AP. Second, readers weighted the use of emotionally-loaded negative language differently.

For both letters, readers noted credibility as closely linked to the evidence thread in AP. In the renewable energy letter, all three readers pointed to the use of evidence as contributing to trustworthiness and as evidence for scoring the letter a 3 for EPV. One explained:

There’s a lot here that establishes [the writer’s] credibility. Their use of the specific details about the oil drilling and how that’s changed over time, gives them a lot of credibility. And then their sourcing of the data from the EPA [Environmental Protection Agency].

These interviewees pointed to the same key words and qualifiers in the CEWAC rubric to support their reasoning for assigning EPV scores. In the “Brutality Towards Police” letter, readers interpreted evidence of credibility differently. Two readers, who scored EPV as a 4, emphasized that this writer’s credibility stemmed from personal experience. One reflected, “While introducing personal data may be ineffective, in this case it’s very effective . . . to convince the intended audience that the issue is important and encourage discourse around the issue.” This reader frames personal experience as data, thus potentially aligning the score with AP’s evidence thread. One reader who rated the paper a 4 on EPV noted that credibility, if scored alone, would have been rated a 3. She explained:

In the anchor papers I’ve seen that were a 4, a lot of credibility . . . was really connected with evidence and how well they were making their point and how well it was backed up. Here, there isn’t really a ton of concrete evidence that I can point to give them this credibility.

The reader who rated EPV as a 2 focused on the relationship between the inclusion of evidence and establishing the credibility of the writer. This reader emphasized, “If you’re looking at this potential topic, that’s full of intensity and emotion . . . , you have to make sure you engage that other side.” The think-aloud interviews echo the findings of the closing survey, in which readers noted the overlap between the credibility strand of EPV and the evidence strand of AP as a concern for attribute independence within a formal, analytic scoring context.

A second issue raised by the interviews is how to weigh emotionally laden language. In contrast to the “Renewable Energy” letter, the “Brutality Towards Police” letter included several examples of emotional language. In considering the use of rhetoric, tone, and style, the three readers who scored the paper a 4 for EPV focused on how the writing “generated empathy” (Interview 4) and “help[ed] the reader identify with the speaker” (Interview 5). They pointed to the writing accomplishing a connection with the audience through acknowledging police violence against Black men and using this rhetorical question, “My question to you Mr. or Mrs. President is what do you plan on doing in order to discourage events like this from happening again?” The two readers who rated EPV lower agreed with their counterparts that personal connection was a strength. However, they focused on the description of the media and people who raised concerns about police misconduct as “ignorant.” While readers who rated EPV as a 4 also noted “ignorant” as potentially problematic, they argued that other aspects of the tone and rhetoric mitigated its use. These different readings of strong, emotional language raise two issues. First, we need to provide additional guidance through anchor papers about how to weigh such language. Second, it points to a larger issue in civic communication. Different audiences and audience members read and respond to language differently, bringing their value systems and cultural/linguistic backgrounds as lenses to their reading; therefore, supporting youth in understanding the potential impact of their words is critical.

ACE. In response to think-aloud interviews and the survey in which 10 respondents (45.5%) called for further clarification, ACE is the attribute we revised most significantly. Table 6 compares the rubric language used during scoring with the revision. The revised language addresses the key challenge expressed by readers in their interviews: how to rate writing that does not include a specific call for action.

Table 6

Comparison of ACE Definition Before and After Pilot Scoring

<u>Score Point</u>	<u>Scoring Conference</u>	<u>Revised</u>
4. The writing	<ul style="list-style-type: none"> • [Call for Civic Engagement] Effectively advocates for increased awareness, a desired change, or direct civic action, in relation to an intended audience. • [Connection] Articulates why the call for civic engagement is an appropriate strategy for addressing the issue. 	<ul style="list-style-type: none"> • [Public Awareness/Importance] Effectively raises awareness about and establishes the public importance of a civic issue in relation to an intended audience. • [Reasonableness/Feasibility] When appropriate, advocates for a desired change or direct civic action, effectively articulating why the action is a reasonable and feasible approach for addressing the issue.

Several readers readily rated a piece of writing regardless of its purpose; for others, writing that only raised awareness served as a point of confusion. A reader articulated this challenge:

To what extent does it advocate for something specific to be done? I'm not seeing that. . . .
 So one of the struggles for me in using the rubric is thinking about . . . in a piece advocating civic engagement, do we need to call for particular actions? Do we need to make suggestions about what the actions should be according to the criteria on the rubric?

In response to this dilemma, the revised language distinguishes writing that raises awareness about a civic issue from that which makes specific proposals for action. For writing that raises awareness, a piece will be rated on how effectively it raises awareness. For writing that proposes action, how reasonable and feasible the action is will distinguish the quality of the writing. How these constructs will be analyzed is the subject of a future study.

AP. Six readers (26.1%) noted on the survey that AP warranted additional clarification. In scoring for AP, all eight interviewees focused on whether the evidence presented was sufficient to support the central argument. This criterion appeared clear and easily recognizable. For the "Brutality Towards Police" letter, all readers recognized that the writing relied almost exclusively on personal experience to support the argument and that personal experience is an acceptable form of evidence. One reader noted, "Personal experience is okay." However, they differed on whether the personal experience offered sufficient support. Similarly, readers consistently applied the alternate views thread for this attribute, readily recognizing alternate views and arriving at general agreement about how effectively they were addressed.

Two issues emerged. First, most interviewees noted they did not understand value structure. Its presence in only the top two score points further muddled the scoring. One reader, for example, wrestled with how to score "Brutality Towards Police" for AP, "I think the preponderance of evidence on the evidence strand and the alternative view strand pulls this paper down toward a 3 in AP, even though

that value structure is so clear and compelling.” The think-aloud interviews, as well as data from the reader survey, prompted us to scale the construct of value structure across all score points.

The second issue centers on reasoning. Three of the eight interviewees didn’t mention reasoning as a rationale for their scores. Four mentioned reasoning but didn’t explain why they assigned a certain score. Only one used an analysis of reasoning to assign a score. This suggests a need for enhanced training focused on reasoning.

ES. For the ES attribute, only readers (13%) noted the attribute would benefit from additional definition. They recommended more genre-specific definitions or more clear anchor papers. Similarly, six of the eight interviewees used the three strands that define the attribute to walk through the structural qualities of the paper and pointed to similar evidence when assigning scores. Overall, ES functioned as intended.

Validity Evidence from External Alignment

To build further validity evidence, we tested its alignment by comparing CEWAC scores for a subsample of papers with scores for the same papers on an external measure of writing, NWP’s Analytic Writing Continuum for Source-Based Argument (AWC–SBA; see Table 7). The AWC–SBA aims to measure general constructs of writing quality, including content, structure, stance, and conventions. An expert AWC–SBA reader scored a subsample of 150 entries from the project Letters to the Next President (See <https://letters2president.org>), which were also scored in the CEWAC scoring session. We expected that CEWAC and AWC–SBA structure-based attributes would have strong alignment. We also posited that CEWAC’s AP attribute may be more highly aligned to the content attribute of AWC–SBA although this alignment was not expected to be perfect given CEWAC’s focus on value structure and use of personal evidence—AWC–SBA only focuses on evidence from print source materials and focuses more on commentary and less on reasoning. The other two CEWAC attributes, EPV and ACE, do not have direct analogs in the AWC–SBA and were not expected to show greater than average alignment.

Table 7

Correlation Matrix of CEWAC and AWC–SBA Attributes

<u>AWC–SBA Attribute</u>	<u>EPV</u>	<u>ACE</u>	<u>AP</u>	<u>ES</u>
SBA_content	0.50	0.55	0.56	0.54
SBA_structure	0.48	0.53	0.56	0.55
SBA_stance	0.45	0.59	0.58	0.55
SBA_conventions	0.38	0.41	0.42	0.42

The lower correlation across the AWC–SBA’s attribute measuring conventions was affirming. Previous studies have found that surface features of writing, such as grammatical errors (Moskal & Leydens, 2000; Rezaei & Lovorn, 2010), can influence readers’ ratings of content dimensions, resulting in construct irrelevant variance (Messick, 1994). By comparing CEWAC scores with AWC–SBA conventions scores, we can see that CEWAC scores are not overly influenced by conventions alone. Further, EPV and ACE demonstrated the average alignment we initially expected. However, the correlation between

structure-based attributes (CEWAC's ES and AWC–SBA's structure) and between CEWAC's AP and AWC–SBA's content were not markedly higher than the other correlations, which differed from our initial expectations. The lack of varied alignment may be due to the high intercorrelation values of the AWC–SBA (Bang, 2013). However, given that our analysis of the independence of CEWAC's attributes could be greater, the lack of varied alignment may further indicate that training and anchor papers should better highlight the independence of CEWAC's attributes.

Discussion

High Quality of CEWAC's Technical Properties

Our analyses showed that CEWAC has a moderate to high level of exact agreement and a high level of exact or adjacent agreement. There is consistency of agreement levels across all attributes. Interrater agreement patterns and agreement patterns with adjudicators were also similar across all attributes. Covariation analyses showed that, even with similar scoring patterns, the attributes held at least a moderate level of independence. This evidence, coupled with robust qualitative evidence around reliability and validity, establishes CEWAC's strong technical properties.

Finalizing the Rubric

The final version of the rubric (Appendix B) reflects shifts in attribute definition and scaling language based on what we learned through the quantitative and qualitative analyses of pilot data and the subsequent review of rubric language. We made revisions to strengthen the independence of attributes by better emphasizing their unique components. Our analyses indicated that attribute wording needed to be improved to create more distinct score points, particularly at the higher end of the scale. Therefore, we made scaling language clearer within attributes and more consistent across attributes.

Potential Limitations

CEWAC has several potential limitations. It is challenging for readers to minimize the impact of their holistic perception of a piece of writing, which is true for all analytic scoring systems. Readers may also use an attribute they feel is most important to weight the others. While some underlying dependence does exist, such as the relationship between a writer's credibility and the way the writer argues the position, it is important for the reader to recognize the independent aspects of each attribute. The final revisions, coupled with strong anchor papers and supporting commentary, will help denote independence. Creating complete objectivity with near perfect agreement is not possible due to the qualitative nature of scoring writing. While scaling language guides the reader, some debate or uncertainty where certain scores should fall for certain papers may always exist, leading to some adjacent scoring.

Civically engaged writing inherently presents a number of other challenges. Both personal experience and emotional language play a large role in civically engaged writing. Readers need to be aware of how those should be considered with each attribute (e.g., how they enhance or reduce credibility, for what purpose the writer may be using these). Civically engaged writing also expresses a wide range of viewpoints around a topic, including varied political ideologies. When readers encounter an argument with which they strongly agree or disagree, they need to be aware of and consider how to counteract those biases while scoring.

Contributions to the Field

CEWAC makes an important contribution to the field of writing assessment by defining and scaling two attributes of particular salience, and sometimes neglected in common rubrics, to public civically engaged writing: EPV and ACE. In addition, this tool explores the relationship between valued academic writing skills—specifically, the ability to develop arguments with sound reasoning and quality evidence—and valued civic skills, which is the ability to contribute thoughtfully to the public sphere. For educators seeking to emphasize writing for authentic public purposes, CEWAC makes visible attributes that are often overlooked in traditional assessment frameworks.

The conceptual challenges with which we continue to wrestle represent challenges in how both adults and youth interact in the civic sphere. They are twofold. First, in public writing, diverse audience members will hear and react to the writer's choices about language, style, and inclusion of personal experience differently; readers' value systems may differ from the writer's. Helping youth understand the potential for varied impact of their words on audiences with different political beliefs, cultural backgrounds, and patterns of language can offer a powerful lens. This is one of CEWAC's inherent tensions. In a formal scoring setting, readers are asked to set aside their perspectives and values to provide ratings of quality that adhere to a set of standards provided by a scoring system. Yet, civic writing is fundamentally about the exchange of ideas, often conflicting ones; readers' and writers' passions, beliefs, and values come together around a piece of writing. As the working group collaborated on CEWAC's development, we frequently returned to this question. Second, CEWAC raises questions central to rhetorical theory since the time of Aristotle's discussion of logos, ethos, and pathos: To what extent can these be disentangled? In the public sphere, an audience's perception about whether a writer's personal qualities (e.g., direct or indirect experience with an issue, common identity with the audience, projection of integrity) qualifies him or her as a trusted source. These questions are important to consider in civic life and teaching as well as in assessment.

Author Note

Linda Friedrich is the director of the Strategic Literacy Initiative at WestEd, San Francisco, CA. She conducted this research as director of Research and Evaluation at the National Writing Project. She holds a PhD in administration and policy analysis from Stanford University's Graduate School of Education and an AB from Bryn Mawr College.

Scott Strother is the lead for Curriculum and Assessment with Carnegie Math Pathways at WestEd. Scott also has held research roles with the Carnegie Foundation for the Advancement of Teaching and the Center for Children and Technology. He has a PhD from the University of Louisville in experimental psychology.

This research was supported in part by grants from the Spencer Foundation and the William and Flora Hewlett Foundation.

Correspondence concerning this article should be addressed to Linda Friedrich, WestEd, San Francisco, CA. lfriedr@wested.org.

References

- Bang, H. J. (2013). Reliability of National Writing Project's Analytic Writing Continuum assessment system. *Journal of Writing Assessment*, 6(1).
- Berry, J. M. (2002). Validity and reliability issues in elite interviewing. *Political Science and Politics*, 35(4), 679–682. doi:10.1017/S1049096502001166
- boyd, d. (2014). *It's complicated: The social lives of networked teens*. New Haven: Yale University Press.
- Civically Engaged Writing Analysis Continuum Working Group. (2016). *Proceedings from the CEWAC working group*. Berkeley, CA: National Writing Project.
- Cohen, C., Kahne, J., Bowyer, B., Middaugh, E., & Rogowski, J. (2012). *Participatory politics: New media and youth political action* (Youth Participatory Politics Survey Project Research Report). Retrieved from https://ypp.dmlcentral.net/sites/default/files/publications/Participatory_Politics_New_Media_and_Youth_Political_Action.2012.pdf
- Culham, R. (2003). *6 + 1 traits of writing*. New York: Scholastic Professional Books.
- DiPardo, A., Storms, B. A., & Selland, M. (2011). Seeing voices: Assessing writerly stance in the NWP Analytic Writing Continuum. *Assessing Writing*, 16, 170–188. doi:10.1016/j.asw.2011.01.003
- Finson, K. D., & Ormsbee, C. K. (1998). Rubrics and their use in inclusive science. *Intervention in School and Clinic*, 34(2), 79–88. doi:10.1177%2F105345129803400203
- Fletcher, J. (2015). *Teaching arguments: Rhetorical comprehension, critique, and response*. Portland, ME: Stenhouse Publishers.
- Gold, D., Garcia, M., & Knutson, A. V. (2019). Going public in an age of digital anxiety: How students negotiate the topoi of online writing environments. *Composition Forum*, 41(Spring 2019).
- Harris, J. (2006). *Rewriting: How to do things with texts*. Logan, UT: Utah State University Press.
- Hess, D. E. (2009). *Controversy in the classroom: The democratic power of discussion*. New York: Routledge.
- Hillocks, G. (2011). *Teaching argument writing, grades 6–12: Supporting claims with relevant evidence and clear reasoning*. Portsmouth, NH: Heinemann.
- Ito, M., Gutierrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salen, K., . . . Watkins, S. C. (2013). *Connected learning: An agenda for research and design*. Irvine, CA: Digital Media and Learning Research Hub.
- Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., & Robinson, A. J. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century* [Occasional paper on digital media and learning]. Chicago, IL: John D. and Catherine T. MacArthur Foundation.
- Kahne, J., Hodgin, E., & Eidman-Aadahl, E. (2016). Redesigning civic education for the digital age: Participatory politics and the pursuit of democratic engagement. The practice of politics has changed. *Theory & Research in Social Education*, 44(1), 1–35. doi:10.1080/00933104.2015.1132646

- Kahne, J., Lee, N., & Feezell, J. T. (2012). Digital media literacy education and online civic and political participation. *International Journal of Communication*, 6, 1–24.
- Kahne, J., & Middaugh, E. (2008). High quality civic education: What is it and who gets it? *Social Education*, 72(1), 34–39.
- Kahne, J., Middaugh, E., & Allen, D. (2014). *Youth, new media, and the rise of participatory politics* (YPP Research Network Working Paper No. 1). Retrieved from https://clalliance.org/wp-content/uploads/files/ypp_workinpapers_paper01_1.pdf.
- Kushin, M. J., & Kitchener, K. (2009). Getting political on social network sites: Exploring online political discourse on Facebook. *First Monday*, 14(11). doi:10.5210/fm.v14i11.2645
- Lazere, D. (2005). *Reading and writing for civic literacy: The critical citizen's guide to argumentative rhetoric*. Boulder, CO: Paradigm Publishers.
- Levine, P. (2008). A public voice for youth: The audience problem in digital media and civic education. In W. L. Bennett (Ed.), *Civic life online: Learning how digital media can engage youth* (pp. 119–138). Cambridge, MA: MIT Press.
- Lynch, D. A., George, D., & Cooper, M. M. (1997). Moments of argument: Agnostic inquiry and confrontational cooperation. *College Composition and Communication*, 48(1), 61–85.
- McWilliams, J. (2013). Lessons from a classroom participatory culture. In H. Jenkins & W. Kelley (Eds.), *Reading in a participatory culture: Remixing Moby-Dick in the English classroom* (p. 161). New York: Teachers College Press.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(13), 13–23. doi:10.3102%2F0013189X023002013
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Mishler, E. G. (1986). *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10).
- National Council for the Social Studies. (2013). *The college, career, and civic life (C3) framework for social studies state standards: Guidance for enhancing the rigor of K–12 civics, economics, geography, and history*. Silver Spring, MD: Author.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Author.
- Parker, W. (2011). Feel free to change your mind: A response to “The potential for deliberative democratic education.” *Democracy and Education* 19(2), 1–4.

- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Purcell, K., Rainie, L., Heaps, A., Buchanan, J., Friedrich, L., Jacklin, A., . . . Zickuhr, K. (2012). *How teens do research in the digital world*. Washington, DC: Pew Research Center's Internet in American Life Project.
- Quinlan, T., Higgins, D., & Wolff, S. (2009). Evaluating the construct-coverage of the E-rater scoring engine. *ETS Research Report Series*, 2009(1), i–35. doi:10.1002/j.2333-8504.2009.tb02158.x
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15, 18– 39.
- Rheingold, H. (2008). Using participatory media and public voice to encourage civic engagement. In W. L. Bennett (Ed.), *Civic life online: Learning how digital media can engage youth* (pp. 97–118). Cambridge, MA: MIT Press.
- Rheingold, H. (2012). *Net smart: How to thrive online*. Cambridge, MA: MIT Press.
- Rubin, B. C. (2012). *Making citizens: Transforming civic learning for diverse social studies classrooms*. New York: Routledge.
- Seidman, I. (2005). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (3rd ed.). New York: Teachers College Press.
- Shresthova, S. (2013). *Between storytelling and surveillance: American Muslim youth negotiate culture, politics and participation* (YPP Research Network Working Paper). Retrieved from <https://ypp.dmlcentral.net/sites/default/files/publications/ShresthovaBetween%20Storytelling%20and%20Surveillance-Working%20Paper%20Report-Sept11-2013.pdf>
- Smith, A. (2013). *Civic engagement and the digital age*. Washington, DC: Pew Research Center's Internet in American Life Project.
- Soep, E. (2014). *Participatory politics: Next-generation tactics to remake public spheres*. Cambridge, MA: MIT Press.
- Sperling, M., & Appleman, D. (2011). Voice in the context of literacy studies. *Reading Research Quarterly*, 46(1), 70–84. doi:10.1598/RRQ.46.1.4
- Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324–328. doi:10.1177%2F003172170508700414
- Swain, S., & Friedrich, L. (2012). *Creating a writing assessment system for research and practice: NWP's Analytic Writing Continuum*. Paper presented at the American Educational Research Association Annual Meeting, Vancouver, BC.
- Swain, S., & LeMahieu, P. (2012). Assessment in a culture of inquiry: The story of National Writing Project's Analytic Writing Continuum. In N. Elliot & L. Perelman (Eds), *Writing assessment in the 21st century: Essays in honor of Edward White* (pp. 45–66). New York: Hampton Press.

Terriquez, V. (2015). Training young activists: Grassroots organizing and youths' civic and political trajectories. *Sociological Perspectives*, 58(2), 223–242. doi:10.1177%2F0731121414556473

Weinstein, E., Rundle, M., & James, C. (2015). A hush falls over the crowd? Diminished online civic expression among young civic actors. *International Journal of Communication*, 9, 84–105.

Wiggins, G. (1994). The constant danger of sacrificing validity to reliability: Making writing assessment serve writers. *Assessing Writing*, 1(1), 129–139. doi:10.1016/1075-2935(94)90008-6

Wolfe, E. W. (2005). Uncovering rater's cognitive processing and focus using think-aloud protocols. *Journal of Writing Assessment*, 2(1), 37–56.

Youniss, J. (2012). How to enrich civic education and sustain democracy. In D. E. Campbell, M. Levinson, & F. M. Hess (Eds.), *Making civics count: Citizenship education for a new generation* (pp. 117–133). Cambridge, MA: Harvard Education Press.

[1] The definition of ACE presented here represents our revision following the pilot scoring conference. The original definition included two threads: call for civic engagement and connection.

Copyright © 2020 - The Journal of Writing Assessment - All Rights Reserved.