

A Persistent Problem in Evaluating Writing

1. For more than a decade standardized testing of English Language Arts (ELA) has included sections on writing. This introduction of writing has been accompanied by rubrics that embody unrealistic expectations about what students are able to write in the limited time available in a testing situation. Nearly ten years ago I wrote the following after conducting research on the ELA test used in New York State:

“Inappropriately inflated rubrics are especially noticeable in tests designed for children. The table below presents a portion of the rubric used to evaluate fourth graders’ writing on the English Language Arts Test in New York State (2002). The first column lists qualities (often referred to as ‘criteria’), while the second provides descriptions of how these qualities are manifested in responses that receive the highest score (level 4).”

Quality	Responses at Level 4
<i>Meaning:</i> The extent to which the response exhibits understanding and interpretation of the task and text(s)	<i>Taken as a whole:</i> <ul style="list-style-type: none">• fulfill all or most requirements of the tasks• address the theme or key elements of the text• show an insightful interpretation of the text• make connections beyond the text
<i>Development:</i> The extent to which ideas are elaborated, using specific and relevant evidence from the text(s)	<i>Taken as a whole:</i> <ul style="list-style-type: none">• develop ideas fully with thorough elaboration• make effective use of relevant and accurate examples from the text
<i>Organization:</i> The extent to which the response exhibits direction, shape, and coherence	<i>The extended response:</i> <ul style="list-style-type: none">• establishes and maintains a clear focus• shows a logical sequence of ideas through the use of appropriate transitions or other devices
<i>Language Use:</i> The extent to which the response reveals an awareness of audience and purpose through effective use of words, sentence structure, and sentence variety	<i>The extended response:</i> <ul style="list-style-type: none">• is fluent and easy to read, with vivid language and a sense of engagement and voice• is stylistically sophisticated, using varied sentence structure and challenging vocabulary

“There is clearly a mismatch between the criteria found in the rubric and the writing that children are able to do in the particular conditions that the test affords. Consider, for example, such descriptions of language use as ‘is fluent and easy to read, with vivid language and a sense of engagement or voice’ and ‘is stylistically sophisticated, using varied sentence structure and challenging vocabulary.’”

“In state education departments throughout the country, phrases like these have been recycled in rubrics used to evaluate what children write on language arts tests. It is disconcerting that standards associated with the highly edited work of seasoned adult writers, working on familiar material over months or even years, is being applied to what children, working under the pressure of a high-stakes test, manage to get on the page when they have about 15 minutes to respond to three tasks about a passage that they have just encountered for the first time.” (Hill, 2004, 1099–1101)ⁱ

ⁱ Hill, C. (2004). Failing to meet the standards: English Language Arts Test for New York State. *Teachers College Record*, 106, 1086–1123.

2. The problem described on the 2002 test is also evidenced on the 2013 test, the first ELA test based on what are called the “common core standards” that have been adopted by 45 states.
- a. Tripartite format used on the 2002 test still used on the 2013 test
 - Section 1: Multiple-Choice Questions
 - Section 2: Constructed-Response Questions (short responses)
 - Section 3: Constructed-Response Questions (extended responses)
 - b. These three sections are now administered over 3 days rather than a single day and the amount of time allotted each day has been increased to 70 minutes. This greater amount of time is, however, paralleled by a greater number of passages (which now tend to be longer) and questions:
 - Section 1: 37 multiple-choice questions based on 6 passages
 - Section 2: 8 short-response questions based on 5 passages
 - Section 3: 2 extended-response questions based on 5 passages
 - c. Separate rubrics are now used to evaluate short responses and extended responses: the rubric used to evaluate extended responses can be found on page 3.ⁱⁱ
3. In “Assessment in the Service of Teaching and Learning,” prepared for the Gordon Commission on the Future of Assessment,ⁱⁱⁱ I proposed an assessment model in which a testing component is complemented by a project component in which student writing is carried out within an extended time frame. Given this extended frame, higher standards can be more realistically brought to bear upon student writing. I illustrate these standards by presenting a rubric that was first developed for the International Baccalaureate to evaluate projects that high school students carried out in the course Theory of Knowledge and then adapted for the Pacesetter Program developed by the College Board to prepare culturally diverse students for higher education.

In developing this rubric, I worked with teachers to identify three areas in which students should concentrate as they learn to evaluate their own writing. As shown in the table below, each area is subdivided to reinforce important values that students need to internalize:

Content	accurate representation of resource material	sufficient support for interpretive statements
Clarity	an overall structure that fits the task	clearly expressed statements in an effective sequence
Critical/Creative Thinking	relevant connections with a larger world	authentic interaction with resource material

The ultimate goal of any assessment of writing is that students internalize basic standards that they can use to evaluate their own work as they move through the educational system and become productive members of society.

ⁱⁱ <http://www.engageny.org/resource/test-guides-for-english-language-arts-and-mathematics>

ⁱⁱⁱ http://www.gordoncommission.org/rsc/pdf/hill_assessment_service_teaching_learning.pdf. This article will appear in a forthcoming issue of the *Teachers College Record* featuring the work of the Gordon Commission that calls for greater integration of assessment with teaching and learning.

New York State Grade 4-5 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0 Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support an analysis of topics or texts	W.2 R.1-9	—clearly introduce a topic in a manner that follows logically from the task and purpose —demonstrate insightful comprehension and analysis of the text(s)	—clearly introduce a topic in a manner that follows from the task and purpose —demonstrate grade-appropriate comprehension and analysis of the text(s)	—introduce a topic in a manner that follows generally from the task and purpose —demonstrate a literal comprehension of the text(s)	—introduce a topic in a manner that does not logically follow from the task and purpose —demonstrate little understanding of the text(s)	—demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.2 W.9 R.1-9	—develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) —sustain the use of varied, relevant evidence	—develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) —sustain the use of relevant evidence, with some lack of variety	—partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant —use relevant evidence inconsistently	—demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	—provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	—exhibit clear, purposeful organization —skillfully link ideas using grade-appropriate words and phrases —use grade-appropriate, stylistically sophisticated language and domain-specific vocabulary —provide a concluding statement that follows clearly from the topic and information presented	—exhibit clear organization —link ideas using grade-appropriate words and phrases —use grade-appropriate precise language and domain-specific vocabulary —provide a concluding statement that follows from the topic and information presented	—exhibit some attempt at organization —inconsistently link ideas using words and phrases —inconsistently use appropriate language and domain-specific vocabulary —provide a concluding statement that follows generally from the topic and information presented	—exhibit little attempt at organization, or attempts to organize are irrelevant to the task —lack the use of linking words and phrases —use language that is imprecise or inappropriate for the text(s) and task —provide a concluding statement that is illogical or unrelated to the topic and information presented	—exhibit no evidence of organization —exhibit no use of linking words and phrases —use language that is predominantly incoherent or copied directly from the text(s) —do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	—demonstrate grade-appropriate command of conventions, with few errors	—demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	—demonstrate emerging command of conventions, with some errors that may hinder comprehension	—demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	—are minimal, making assessment of conventions unreliable

*logically
insightful*

grade-appropriate

*generally
a literal*

*not logically
little*