

Eric D. Turley and Chris W. Gallagher

On the *Uses* of Rubrics: Reframing the Great Rubric Debate

Reframing the Great Rubric Debate beyond the taking of sides, Eric D. Turley and Chris W. Gallagher propose a set of questions designed to help educators assess the value of rubrics. They contend that these four simple questions push educators to deliberately articulate “judgments about the *uses* of rubrics” and why they will or will not use them for assignments.

In the pages of *English Journal* and beyond, the Great Rubric Debate is on. Rubrics are sorting machines—or they are useful instructional tools. Rubrics lead to standardization—or they help us connect with individual students. Rubrics rest on false claims to objectivity—or they make subjectivity visible. Rubrics put teachers and students on autopilot—or they enrich conversations between teachers and students. It is time to move beyond rubrics—or it is time to embrace them. The titles and subtitles of articles and books about rubrics are either dismissive or defensive: *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing* (Broad), *Rethinking Rubrics in Writing Assessment* (Wilson), “The Trouble with Rubrics” (Kohn), “In Defense of Rubrics” (Spandel), “Why I Won’t Be Using Rubrics to Respond to Students’ Writing” (Wilson).

Notice the common denominator in these pieces: the acknowledgment that a rubric is a *tool*. As high school and college writing teachers, we have learned a great deal from the authors cited here. However, we believe the debate has been framed in limited ways. It makes little sense to dismiss or embrace a tool as such—that is, without careful attention to why, how, by whom, and in what contexts it is used. If the Great Rubric Debate remains focused on whether rubrics are “good” or “bad,” it will be useful only to the extent that it allows us to take sides—an emotionally satisfying result, perhaps, but not an educationally meaningful one. Our aim, then, is to offer a reframing of the

debate that allows us to explore the benefits, limitations, and dangers of rubrics.

Specifically, we propose a set of questions that we believe can help us assess the value of rubrics (or any pedagogical tool):

1. What is the tool for?
2. In what context is it used?
3. Who decides?
4. What ideological agenda drives those decisions?

These questions provide a heuristic for making judgments about the *uses* of rubrics.

We want to be clear at the outset that we are *not* saying rubrics are “merely” neutral tools. On the contrary, our heuristic amplifies the important critiques made by critics of rubrics by forcing us to attend to their users and uses in various contexts. It also asks us to articulate the conditions, if any, under which rubrics might be used appropriately. Individuals will answer the questions differently, but the heuristic forces us to articulate the terms on which we are willing to use, or refuse to use, this tool.

If anything, our heuristic is skeptical: It cautions us to be deliberate, careful. Pedagogically, this seems wise, especially when we consider what

If the Great Rubric Debate remains focused on whether rubrics are “good” or “bad,” it will be useful only to the extent that it allows us to take sides—an emotionally satisfying result, perhaps, but not an educationally meaningful one.

we think of as the “law of distal diminishment.” The law of distal diminishment says that any educational tool becomes less instructionally useful—and more potentially damaging to educational integrity—the farther away from the classroom it originates or travels.

The law of distal diminishment becomes clear when we examine several uses of rubrics. We begin with a brief exploration of the historical origins of rubrics as writing scales. We then turn our attention to contemporary state-imposed and classroom rubrics.

The Origins of Rubrics

In November 1912, Ernest C. Noyes argued in *English Journal* that a “new science of education based upon exact measurement and judgment by ascertained facts” was needed in composition (532). Noyes advocated standardized measurement of composition because “[o]ur present methods of measuring compositions are controlled too much by personal opinion, which varies with the individual. What is wanted is a clear-cut, concrete standard of measurement which will mean the same thing to all people in all places and is not dependent upon the opinion of any individual” (534). Noyes used a thermometer to exemplify a tool that produced a scientific standard; composition needed a scale that would allow the measurement of writing to be as universal as a Fahrenheit scale makes the measurement of temperature.

The creation of writing scales became a popular endeavor over the next decade as university professors designed and calibrated scales that could measure “good” composition in schools across the nation.¹ The original writing scale, *A Scale for the Measurement of Quality in English Composition by Young People*, was more commonly referred to as the Hillegas Scale after its architect, Milo B. Hillegas, a professor at Columbia University’s Teachers College. Hillegas, a student of Edward Thorndike’s, advanced Thorndike’s attempt to quantify learning through positivist measurement. According to Thorndike, “Whatever exists, exists in some amount. To measure it is simply to know its varying amounts” (379). In his creation of a writing scale, Hillegas offered a scientific way to quantify the quality of student compositions fashioned on a statistical model of normal distribution.²

The Hillegas Scale was adopted in many schools. The teacher was to compare a student’s writing from his or her class with the samples on the scale and record the numerical ranking of the student’s writing. This practice codified a belief that students’ writing could be translated into a number and quantified. The Hillegas Scale and the others that followed answered the call for a scientific tool that could eliminate teacher subjectivity for an objective and exact numerical measurement of student writing.

It is important to note that these scales were not intended to improve student writing (Hudelson 163). Rather, the architects of the scales and the school administrators who adopted them had other purposes in mind. Hillegas intended his scale to provide a standardized form of measurement that would allow administrators and investigators to “measure and express the efficiency of a school system” so that comparisons and rankings could be made between schools across the nation (2). C. C. Certain demonstrated how writing scales could be used to compare schools in his 1923 *EJ* article “Are Your Pupils Up to Standard in Composition?”³ As part of his study, Certain gave the students in X-Town a “standard test” in narrative writing—students had forty-five minutes to write to the prompt of “An Experience Which I Shall Never Forget.” Certain compared thirteen cities across the nation in eighth-grade composition. He ranked the median scores on the test and argued that the results of the composition test in X-Town High School were inferior to those of the other high schools where comparable data were available (368).

While investigators were interested in comparing schools, school principals used writing scales to measure the effectiveness of their teachers. S. A. Courtis, supervisor of educational research in Detroit, Michigan, was particularly interested in using writing scales as a way to measure effective teaching. “The Hillegas scale,” Courtis proclaimed, “is of value in determining the efficiency of different methods of teaching” (213). As Courtis wrote, “I believe every teacher should compare both the amount of the change she has produced during the term with the standard change produced by other teachers of her grade for the city as a whole, and the final ability of her class with the standard for the city as a whole” (215). By requiring teachers to use

the Hillegas Scale in their classrooms, Curtis was able to measure their performance through the increase or decrease of student writing scores as they compared to those in other schools. If teachers were able to maintain student scores at an acceptable level, Curtis left them alone. However, if their scores dropped, the teacher and administrator would work to “bring her work up to standard” (215). If the teacher could not obtain the required results, she risked being dismissed.

This brief history of writing scales provides answers to the four questions we posed above. Early writing scales were constructed to provide an objective numerical measurement of student writing in order to allow comparisons among students, teachers, classes, schools, cities, and so on. The scales were created outside of the schools, typically at universities, and then implemented by school administrators. Teachers and students had no hand in creating the scales and were not expected to use them to improve student writing. The ideology that underwrote writing scales hinged on managerial efficiency and positivist science, including the idea that technological precision would replace subjective human judgment.

Today, state-imposed rubrics are still created and implemented to generate numbers for accountability rather than improve student writing. Standardized writing assessment at the elementary and secondary school levels requires students to create an essay within a specific mode (narrative, expository/descriptive, persuasive), and their sample is scored against a state rubric that most often is based on 6-trait or 6-trait+1 writing rubrics. Many teachers, attempting to make the best of these rubrics, use them in their classrooms to provide a “common language” for describing “good writing.” While understandable, this practice has limited utility for improving writing because it uses a predetermined set of criteria that teachers and students had no hand in creating. Our contention—based on the law of distal diminishment—is that because imposed rubrics originate outside of classrooms, their primary function will be to constrict and constrain writing. These rubrics “improve” writing only to the extent that we define “improvement” as conforming to a set of imposed expectations.

Ultimately, state rubrics represent the consummation of the positivistic dream of a science of

education that would overcome teachers’ subjectivity in the name of precision and efficiency. Their provenance should encourage us to view rubrics skeptically. However, as we have suggested, not all rubrics are born equal; there are different kinds, and different uses, of rubrics.

Classroom Rubrics

Some rubrics originate in classrooms and never travel beyond those four walls. Can these rubrics be salvaged even within constructivist classrooms, where the goal is to help students come to their own meanings and their own purposes for writing? Or are rubrics, as their critics claim, trapped within the positivist/behaviorist ideological framework in which they were conceived?

To be constructivist, rubrics must not require standardization. To some critics of rubrics, this will sound like a contradiction; they believe rubrics necessarily require standardization. Certainly rubrics *do* have this effect when they are imposed on classrooms from outside; in this case, student writing becomes an exercise in target practice. But rubrics can also be descriptive, rather than prescriptive. After all, rubrics are nothing more or less than *classification devices*: tools that allow us to arrive at and enunciate judgments—which we and our students do all the time.

Much of the time, we don’t need classification devices; it is enough simply to converse with individual students about their texts, to respond as readers, and as Maja Wilson suggests, “to look to the piece of writing itself to suggest its own evaluative criteria” (*Rethinking* 42). But in our high school and college classrooms, we also want to create *writing communities* in which we develop a shared vocabulary for talking about—and rendering judgments about—writing. This is important to us at grade time, to be sure, but also at less formal moments throughout our courses.

In some classes, we find it useful to use rubrics to assist us and students in arriving at and

Early writing scales were constructed to provide an objective numerical measurement of student writing in order to allow comparisons among students, teachers, classes, schools, cities, and so on. The scales were created outside of the schools, typically at universities, and then implemented by school administrators.

rendering these judgments. It is helpful to ask students to consider how well they have accomplished the goals they set both individually and as a community of writers. Categories of performance can be named in myriad ways beyond single numbers or letters; we can use adjectives or adjectival phrases (one of Chris's classes used the following descriptors: Totally Rocks, Rocks Quietly, Sort of Has a Beat, and Does Not Rock), or we can avoid using any overall descriptors and instead focus on specific indicators. Here, because we imagine grading rubrics are the most common form of classroom rubrics, we focus on an example of that type.

See Figure 1 for a classroom rubric from one of Chris's first-year college writing classes.

This rubric is different from the state-imposed rubrics described above. It is messier and more context-specific; indeed, we hesitated to include it at all for fear it would be unintelligible to readers who were not a part of the class that produced it. That is our point: This rubric is a product of a particular community of writers. It is also the product of a particular moment in time; it is, in fact, a midterm portfolio rubric. In addition to providing a tool for generating midterm "grades-so-far," as Chris calls them, the rubric also served as a tool for Chris and his students to set goals for the second half of the semester. At the end of the semester, the class revisited and revised the rubric—drastically. Even within the context of this course, the rubric was ephemeral: a tool judiciously used and then put back down.

However modest, we believe rubrics like this can be important tools for building constructivist classrooms. This tool was used constructively in several ways:

- > It was created *with* students and reflects their values, goals, and language. For instance, the phrase "put readers in the moment" is a bit of insider language that this group of writers found useful. While this is an unconventional way to describe "voice," *these* writers knew what it meant and wanted to hold themselves to that aspiration. Our negotiations with students have led us to use words that students value more highly than we do—one of their favorites is "flow"—and to put more emphasis on some features of writing than we otherwise would have

because we want to honor what *this* community of writers values.

- > It was created only *after* students drafted. For a rubric to function descriptively, rather than prescriptively, it must follow, rather than precede, students' initial attempts. Once the rubric is generated, it does guide students' work—as it should, because it represents what the students themselves say they value about writing they have done and are trying to do. Even at this stage, we encourage students to use rubrics as guides, not blueprints.
- > Students used it to arrive at and render judgments about their own and each others' writing. The similarities and differences between and among the judgments of Chris and the students became a rich source of conversation. In our experience, the most interesting and useful exchanges occur around *differing* evaluations. In these cases, we never want to sacrifice rich conversation for reader agreement, as so often happens when rubrics are used for accountability purposes; instead, we work with writers on making sense of and responding to multiple readings—work that all writers must learn to do. (We have also found ourselves shifting our judgments in the face of conflicting readings.)
- > It is flexible enough to include an additional criterion (the "wild card") that students believe is important to consider for their particular portfolio. While most goals that individuals set for their writing overlap with those that the community of writers deems important, individual writers often have additional or different goals as well. The "wild card" offers students the opportunity to identify one or more of those goals and to evaluate their work accordingly. In addition, we encourage students to use author's notes and portfolio cover letters to discuss particularities of their work that fall outside of the rubric.
- > It includes a provision for narrative comments. Even though this rubric is a grading rubric, it calls for a discursive rendering of judgments. To be sure, these comments sometimes tend to justify the grade, as is often the case in any grading context. But more often, in our experience, they offer rich fodder for conversations with students.

FIGURE 1. Classroom Rubric

	A	B	C	D	F	NOTES
Purposes and Audiences <i>Writer shows...</i>	consistent awareness of purposes and audiences	consistent awareness of purposes and audiences	mostly consistent awareness of purposes and audiences	inconsistent awareness of purposes and audiences	little awareness of purposes and audiences	
Voices <i>Writer...</i>	consistently puts readers in the moment; uses voice to make portfolio rich, interesting	usually puts readers in the moment; uses voice to make portfolio rich, interesting	sometimes puts readers in the moment; uses voice to make portfolio rich, interesting	does not generally put readers in the moment	does not put readers in the moment	
Writing Strategies <i>Writer...</i>	understands, effectively uses variety of well-chosen writing strategies	understands, effectively uses variety of mostly well-chosen writing strategies	understands, uses some writing strategies; not always well-chosen or well executed	often uses ill-chosen or poorly executed writing strategies	uses ill-chosen or poorly executed writing strategies	
Conventions <i>Final drafts...</i>	demonstrate mastery of conventions; virtually error-free	demonstrate control of conventions; few, unobtrusive errors	demonstrate uneven control of conventions; several unobtrusive errors	demonstrate lack of control of conventions; some obtrusive errors	demonstrate lack of control of conventions; many obtrusive errors	
Reflection <i>Evidence of...</i>	consistent, insightful reflection for each project and across projects	mostly consistent, insightful reflection for each project and across projects	significant, if sometimes inconsistent or perfunctory, reflection	inconsistent or perfunctory reflection	little or no reflection	
Revision <i>Evidence of...</i>	consistent, effective revision in and across projects	mostly consistent, effective revision in and across projects	sometimes inconsistent or perfunctory revision in and across projects	inconsistent or perfunctory revision	little or no revision	
Organization <i>Evidence of...</i>	careful, thoughtful organization of portfolio, projects	careful, thoughtful organization of portfolio, projects	mostly effective organization of portfolio, projects	ineffective organization of portfolio, projects	little or no attention to organization of portfolio, projects	
Wild Card						

In short, this rubric does not replace engaged response; it is a tool for generating more of it. While the language of the rubric represents a consensus of the values of this community of writers, it is also a launching point for conversation: a place to start, and never a place to end.

Conclusion

Rubrics are created for many purposes in many contexts, including classrooms, programs, institutions, and states. Instead of declaring all rubrics “good” or “bad,” we need to examine what they do, why, and in whose interests. Rubrics designed by teams of teachers for programmatic assessment, for instance, might be good at helping teachers articulate—and thereby clarify—their values and identify trends in student writing, and perhaps also to help administrators to know where to target professional development resources. But that same rubric might then be used to challenge teachers’ grading practices—a use that problematically conflates (or confuses) programmatic assessment with classroom assessment. This move takes a tool that is perfectly valid in one context—a program—and warps it by wielding it for another purpose in another context—a classroom. This is why our four-question heuristic is so important: It bids us to deliberate not only on the tool itself but also on how, why, and by whom it is being used. Like a rubric, our heuristic is a classification device that helps us render informed judgments.

In reframing the debate on rubrics and providing our four-question heuristic, we hope to offer teachers ways of identifying uses and abuses of rubrics in different contexts. If as teachers we are handed a rubric—from the state, district, or the teacher next door—we need to consider the law of distal diminishment and be skeptical of the ability of that rubric to improve students’ writing. In addition, the four-question heuristic is an attempt to keep those involved at different levels of assessment honest. Any time a rubric is used, its proponents—whether teachers, university faculty, policymakers,

school administrators, or private vendors—ought to be able to answer sufficiently the four questions we pose. Ideally, posing these questions would lead to only sparing and judicious uses of rubrics. At the least, it would lead us into a debate worth having.

Notes

1. A decade after Noyes’s call to create composition scales, several scales were being used in schools and discussed in journals. In his 1923 article, Hudelson cites scales designed by M. B. Hillegas, E. L. Thorndike, M. R. Trabue, E. E. Lewis, F. W. Ballou, F. S. Breed and F. W. Frostic, M. H. Willing, and M. J. Van Wagenen (164–67).

2. Starting with 7,000 samples of writing, Hillegas pulled out 83 samples that demonstrated a range in writing by young people. He had over 400 judges read and rank the papers. After creating a zero point, Hillegas set the median deviation at 75 percent. If 75 percent of the judges deemed the quality of one paper stronger than another, it was placed one deviation unit apart on the scale. The final scale consisted of ten samples of writing plotted numerically from 0 to 1,000.

3. Certain served as the chair of NCTE’s Committee on Examinations. *English Journal* ran a series of articles in 1923 encouraging teachers and supervisors to incorporate “standard tests” in composition in their schools.

Works Cited

- Broad, Bob. *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*. Logan: Utah State UP, 2003.
- Certain, C. C. “Are Your Pupils Up to Standard in Composition?” *English Journal* 12.6 (1923): 365–77.
- Courtis, S. A. “The Value of Measurements: The Uses of the Hillegas Scale.” *English Journal* 8.4 (1919): 208–17.
- Hillegas, Milo B. *A Scale for the Measurement of Quality in English Composition by Young People*. New York: Teachers College, 1913.
- Hudelson, Earl. “The Development and Comparative Values of Composition Scales.” *English Journal* 12.3 (1923): 163–68.
- Kohn, Alfie. “The Trouble with Rubrics.” *English Journal* 95.4 (2006): 12–15.
- Noyes, Ernest C. “Progress in Standardizing the Measurement of Composition.” *English Journal* 1.9 (1912): 532–36.
- Spandel, Vicki. “In Defense of Rubrics.” *English Journal* 96.1 (2006): 19–22.
- Thorndike, Edward L. “Measurement in Education.” *Teachers College Record* 22.5 (1921): 371–79.
- Wilson, Maja. *Rethinking Rubrics in Writing Assessment*. Portsmouth: Heinemann, 2006.
- . “Why I Won’t Be Using Rubrics to Respond to Students’ Writing” *English Journal* 96.4 (2007): 62–66.

Eric D. Turley, a former high school teacher, is a PhD candidate in English at the University of Nebraska–Lincoln. His research interests include writing assessment, politics of literacy, and school reform. *email*: eturley1@bigred.unl.edu. **Chris W. Gallagher** is associate professor of English at the University of Nebraska–Lincoln, where he teaches courses in writing and teaching. He is the author of *Radical Departures: Composition and Progressive Pedagogy* (NCTE, 2002) and *Reclaiming Assessment: A Better Alternative to the Accountability Agenda* (Heinemann, 2007), as well as numerous articles. *email*: cgallagher2@unl.edu.