The Trouble with Rubrics (#)
The Trouble with Rubrics

By Alifie Kohn

Once upon a time I vaguely thought of assessment in dichotomous terms: The old approach, which consisted mostly of letter grades, was crude and uninformative, while the new approach, which included things like portfolios and rubrics, was detailed and authentic. Only much later did I look more carefully at the individual floaters rolling by in the alternative assessment parade and stop cheering.

For starters, I realized that it’s hardly sufficient to recommend a given approach on the basis of its better being than old-fashioned report cards. By that criterion, just about anything would seem preferable to what got us to the point of rubrics in the first place. Rubrics come in all shapes and sizes, but not all alternative assessment rubrics in the same genre. My growth came from realizing that the assumptions on which this technique rested and also the criteria by which they (and assessment itself) were typically judged were deeply flawed. Doubts were stoked not only by murmurs of dissent from thoughtful educators but by the case made for the technique by its enthusiastic proponents. For example, I read in one article that “rubrics make assessing student work quick and efficient,” and they help teachers to justify to parents and others the grades they assign to students.”[2] To which the only appropriate response is: oh—first of all, something that’s pitched to teachers as a handy strategy of self-justification during parent conferences “Look at all these 3’s, Mrs. Gromow! How could I have given a B?” doesn’t seem particularly promising for inspiring teachers to improve their practices, let alone rethink their premises. Finally, I’m under the impression that promise to make our pedagogical papers “quick and efficient” is also a bit of an appeal here, but the best teachers would react to that selling point with skepticism, if not disdain. They’d immediately ask what we had to sacrifice in order to spit out a series of tidy judgments about the quality of student learning. To ponder that question is to understand how something that presents itself as an innocuous scoring guide can be so profoundly wrongheaded.

Consistent and uniform standards are admirable, and maybe even workable, when we’re talking about, say, the manufacture of DVD players. The process of trying to gauge children’s understanding of ideas is very different, however. It necessarily entails the exercise of a human judgment, which is an imprecise, subjective affair. Rubrics are, above all, a tool to promote standardization, to turn teachers into grading machines or at least allow them to pretend that what they’re doing is exact and objective. Frankly, I’m amazed by the number of educators whose opposition to standardized tests and standardized curricula mysteriously fails to extend to standardized in-class assessments.

The appeal of rubrics is supposed to be their high interrater reliability, finally delivered to language arts. A list of criteria for what should be awarded the highest possible score when evaluating an essay is supposed to reflect near-unanimous taste. Even more, it’s supposed to ensure that students are graded fairly and that they tend to think less deeply, avoid taking risks, and lose interest in the learning itself.[2] The ultimate goal of authentic assessment must be the elimination of grades. But rubrics actually help to turn a classroom into a test-preparation factory, so it’s possible to get a bunch of people to agree on what rating to give an assignment as long as they’re willing to accept and apply someone else’s narrow criteria for what merits that rating. Once we check our judgment at the door, we can all learn to give a 4 to exactly the same things. This attempt to deny the subjectivity of human judgment is objectionable in its own right. But it’s also harmful in a very practical sense. In an important article published in 1999, Linda Mabry, now at Washington State University, pointed out that rubrics “are designed to function as scoring guidelines, but they also serve as arbiters of quality and agents of control over what is taught and valued. Because agreement among scorers is more easily achieved with regard to such matters as spelling and organization,” these are the characteristics that will likely find favor in a risk-averse classroom. Rubrics signify approaches that “compliance with the rubric tended to yield higher scores but produced ‘mechanical’ writing.”[3]

To this point, my objections assume only that teachers rely on rubrics to standardize the way they think about student assignments. Despite my misgivings, I can imagine a scenario where teachers benefit from consulting a rubric briefly in order to think about various aspects of a student’s work. But the rubric will turn into an unthinking tool for grading if teachers then turn the class into a test-preparation factory, so it’s possible to get a bunch of people to agree on what rating to give an assignment as long as they’re willing to accept and apply someone else’s narrow criteria for what merits that rating. Once we check our judgment at the door, we can all learn to give a 4 to exactly the same things. This attempt to deny the subjectivity of human judgment is objectionable in its own right. But it’s also harmful in a very practical sense. In an important article published in 1999, Linda Mabry, now at Washington State University, pointed out that rubrics “are designed to function as scoring guidelines, but they also serve as arbiters of quality and agents of control over what is taught and valued. Because agreement among scorers is more easily achieved with regard to such matters as spelling and organization,” these are the characteristics that will likely find favor in a risk-averse classroom. Rubrics signify approaches that “compliance with the rubric tended to yield higher scores but produced ‘mechanical’ writing.”[3]

This is the sort of outcome that may not be noticed by an assessment specialist who is essentially a technician, in search of practices that yield data in ever-greater quantities. “A B+ doesn’t mean anything; it’s just a grade. A B+ means ‘OK’ whatever that means. There’s nothing there to discover rather than to decide” which essays meet those criteria.

Now some observers criticize rubrics because they can never deliver the promised precision; judgments ultimately turn on adjectives that are murky and up being left to the teacher’s discretion. But all bets are off if student responses are graded and assessed by standards that rely on the judgment of others. “Standardized tests scores as an assessment of the quality of teaching is a
disaster.” Wendy Wilson, a Michigan teacher and former school board member told me that she began “resisting the rubric temptation” the day “one particularly uninterested student raised his hand and asked if I was going to lose interest in the learning itself.”[2] The ultimate goal of authentic assessment must be the elimination of grades. But rubrics actually help to turn a classroom into a test-preparation factory, so it’s possible to get a bunch of people to agree on what rating to give an assignment as long as they’re willing to accept and apply someone else’s narrow criteria for what merits that rating. Once we check our judgment at the door, we can all learn to give a 4 to exactly the same things. This attempt to deny the subjectivity of human judgment is objectionable in its own right. But it’s also harmful in a very practical sense.

The appeal of rubrics is supposed to be their high interrater reliability, finally delivered to language arts. A list of criteria for what should be awarded the highest possible score when evaluating an essay is supposed to reflect near-unanimous taste. Even more, it’s supposed to ensure that students are graded fairly and that they tend to think less deeply, avoid taking risks, and lose interest in the learning itself.[2] The ultimate goal of authentic assessment must be the elimination of grades. But rubrics actually help to turn a classroom into a test-preparation factory, so it’s possible to get a bunch of people to agree on what rating to give an assignment as long as they’re willing to accept and apply someone else’s narrow criteria for what merits that rating. Once we check our judgment at the door, we can all learn to give a 4 to exactly the same things. This attempt to deny the subjectivity of human judgment is objectionable in its own right. But it’s also harmful in a very practical sense.

This attempt to deny the subjectivity of human judgment is objectionable in its own right. But it’s also harmful in a very practical sense.

The appeal of rubrics is supposed to be their high interrater reliability, finally delivered to language arts. A list of criteria for what should be awarded the highest possible score when evaluating an essay is supposed to reflect near-unanimous taste. Even more, it’s supposed to ensure that students are graded fairly and that they tend to think less deeply, avoid taking risks, and lose interest in the learning itself.[2] The ultimate goal of authentic assessment must be the elimination of grades. But rubrics actually help to turn a classroom into a test-preparation factory, so it’s possible to get a bunch of people to agree on what rating to give an assignment as long as they’re willing to accept and apply someone else’s narrow criteria for what merits that rating. Once we check our judgment at the door, we can all learn to give a 4 to exactly the same things. This attempt to deny the subjectivity of human judgment is objectionable in its own right. But it’s also harmful in a very practical sense.