

## Formative and Punitive Assessment

Randall Curren  
*University of Rochester*

Matthew Hayden takes aim at high stakes uses of student assessments to evaluate, reward, and penalize teachers. He argues convincingly that such uses of assessments are misguided and counterproductive. In making that argument, he relies on aspects of my analysis in “Academic Standards and Constitutive Luck,”<sup>1</sup> evidence that teachers’ efforts account for only three percent of measurable variation in student achievement, and the research of my colleagues, Richard Ryan and Edward Deci. His essay may be regarded as a friendly amendment or extension of my own, and I am sympathetic to his main conclusions. We agree that students and teachers should not be judged on the basis of results they cannot control, we agree that grades should be used to promote learning and fair equality of opportunity, and we agree that the information provided by assessment of student learning should be used formatively to promote good teaching — and rarely, if ever, to sanction teachers without more direct knowledge of the quality of their teaching. The logic by which Hayden extrapolates from fairness-to-students to fairness-to-teachers is not as straightforward as it seems, however.

I should begin by noting that Hayden’s description of my paper as particularly concerned with Outcomes Assessment (OA) is inaccurate. The paper makes some limited claims about Outcomes Assessment in the context of addressing features of traditional classroom evaluation practices. While Hayden says that “OA could be any tool used by a teacher to assess student learning,” I have in mind OA *systems* that aim to ensure there is articulated alignment between evaluation, instruction, and curriculum. In some higher education contexts, OA is not much more than an exercise that instructors engage in to ensure they and their students are clear about the character of that alignment. In other contexts OA is far more demanding and may compel a realignment of instruction with assessments of minimal competence. OA should not be equated with ordinary classroom practice, nor should it be equated with high stakes testing and accountability schemes.

Hayden suggests “we need to think more carefully about” the use of outcome measures “as summative tools for teacher effectiveness in the same way that Curren suggests caution regarding student achievement and OA.” My cautionary lesson concerns meritocratic uses of traditional grade records, however, not OA. The distinction between *formative* and *summative* evaluation is also less helpful in framing Hayden’s diverse arguments than a richer set of distinctions would be.

The *normatively* salient distinction suggested by my own work is between *formative* and *punitive* evaluation — evaluation pursuant to withholding benefits or imposing penalties. The use of outcome measures of student learning to motivate better teaching and to sanction teachers and schools whose students do not perform well reflects a broader tendency to overestimate the efficacy and legitimacy of *force* — a cultural tendency to divide the world into those who choose to be good and those

who choose to be bad, ignoring the realities of human development, capacity-building, motivation, and context.

The research of Deci, Ryan, and their collaborators is important in establishing that the *motivational* effects of high stakes tests are not what is often supposed — they induce anxiety, causing teachers to be more controlling and less effective. A motivationally salient contrast could be drawn between *outcome* and *activity* measures. Ryan has aptly observed that the behaviorist orientation of high stakes testing and accountability schemes would commend reinforcement of teacher *behaviors* or *activities*, not learning *outcomes*.<sup>2</sup> Hayden is right to point out that student outcomes are only marginally attributable to teaching practices, and he is right to commend the NBCT (National Board Certified Teachers) criteria for certification, which focus much more on activities than on outcomes. The *outcome-activity* contrast captures an important dimension of this, while also providing a useful perspective on the incentives to cheat, game the system, and teach to the test generated by reinforcement not of good teaching practices but of measured outcomes.

To *normative* and *motivational* perspectives on evaluation, we should add a *diagnostic* perspective. Hayden recommends a “focus on the practices or process[es] that produce” outcomes so we can understand “*what* specifically has gone wrong and why.” I agree. To the extent that students succeed through the efforts of teachers, our diagnostic evaluative focus should be on teacher *capabilities* and their fulfillment in teaching *practices* or *activities* (the NBCT criteria, minus outcomes), and on equipping teachers with diagnostically useful evaluations of students. *Process measures* that reveal how students think may be defended on both diagnostic and normative grounds.<sup>3</sup>

The argument by which Hayden has extrapolated from my position on student evaluation faces obstacles, but the distinctions I have outlined suggest a different way of organizing his most potent arguments. He is right that what constitutive luck (luck in how a person is constituted) is from a student’s perspective has a bearing on what a teacher can be responsible *for*. From the teacher’s perspective this is a matter of *resultant luck*, or luck in the way things turn out. Hayden says the strategy of his essay is to apply my “caution regarding student achievement and OA” arising from constitutive luck “to the same conditions for teachers.” He goes on to say that my “analysis recognizes that students ... should not be held accountable for their academic performances ... at least until they have reached an ‘age of reason.’” This is a bit complicated.

Philosophers have distinguished four kinds of moral luck and they disagree about whether they are real and, if they are real, what they imply. I fall into the camp of those who deny the reality of *resultant luck*, affirm the existence of *constitutive luck*, and argue that *constitutive luck* is compatible with the forms of control required for responsibility. This means that I would join Hayden in insisting that teachers are not responsible for the deficiencies of student learning they cannot control, but contrary to what he suggests, I cannot (and do not) say that students “should not be held responsible for their academic performances.” I say that “assigning a grade is

tantamount to holding a student responsible for an academic performance,” and that constitutive luck “doesn’t seem to” matter “to the fairness of grades.”<sup>4</sup> My objection to using grades as external credentials pertains to fair equality of opportunity, not constitutive luck. Fair opportunity pertains to the qualifications for a position and the opportunity to acquire those qualifications. In this respect, fairness to students and fairness to teachers share common features, but those common features do not fully justify Hayden’s counterparts to the first two “suggestions for assessment reform” he rightly attributes to me: his claims that assessment of teachers should aim to equalize teaching attainment, and that only after it has done that for some predetermined period of time can it be legitimately used in personnel decisions.

Let us suppose that fair equality of opportunity requires not only education through high school that provides equal prospects of developing talents and job qualifications, but also “equal *teaching* attainment” or a world in which teachers are equally proficient. Because teachers are adults whose developed talents are properly considered in employment decisions, not children who deserve equal chances to develop their talents, it would not follow that test scores should not be used as credentials. Nor would it follow that some waiting period is required *as a matter of justice* before such scores could be used in personnel decisions.

Different considerations must be adduced to defend these conclusions, considerations pertaining to aspects of educational systems not addressed by the ideal of fair opportunity. I have a suggestion. The phenomenon of early career teacher attrition suggests it would be more productive and humane to focus a greater proportion of investments in good teaching on the work place and a smaller proportion on pre-service teacher education. Structuring investments in this way would be less wasteful and would subject aspiring teachers to less *risk* that their investments in becoming teachers will go to waste. Early career punitive outcome evaluations of teachers arguably make an existing problem worse — a problem of ensuring there are enough good teachers and distributing the associated risks humanely.

The problem with *punitive outcome evaluation* of teachers is in part that measurable student learning *outcomes* are a very poor measure of teacher *quality* and meaningful measures of quality are essential to *fair opportunity*, in part that it is *motivationally* misguided and counterproductive, and in part that it would be more productive and humane to focus resources on *diagnostic* evaluation to promote good teaching.

---

1. Randall Curren, “Academic Standards and Constitutive Luck,” in *A Teacher’s Life: Essays for Steven M. Cahn*, eds. Robert Talisse and Maureen Eckert (Lanham, MD: Lexington Books, 2009), 13–32.

2. Richard Ryan, personal communication with the author.

3. See Randall Curren, “Coercion and the Ethics of Grading and Testing,” *Educational Theory* 45, no. 3 (1995): 425–41.

4. Curren, “Academic Standards and Constitutive Luck,” 16, 18.