

CHAPTER 2.

HOW I CAME TO LABOR-BASED GRADING CONTRACTS

To a dyed-in-the-wool grader, the thought that grades are disturbing can come as a surprise. That grades become badges of autocratic or despotic judgments of students has not bothered them. That grades introduce a serious pedagogic fault, with a marked and often adverse influence on teaching itself, does not occur to them.

A friend of mine who taught dental students favored the use of grades but was willing to weigh the question. He once lined up some large models of teeth and asked me to look at them. He pointed out the smooth contours, perfect fits, and strength of fillings in the teeth at the beginning of the row and then described carefully the progressive irregularities in teeth farther down the row. Having told me about the differences, he pointed out that I could see them clearly, though I was not a dentist. Then he asked triumphantly, "Why shouldn't I grade them from A to F?" My answer startled him. I merely said, "You just told me why you should not." He looked puzzled for a moment and then began to smile. He had been telling me specific features that he thought were right or needed correction. Not his judgment of rank but his specific observations were pertinent, whether he was teaching or just talking to me. Imagine that you are a prospective employer, and I know a student better than you do. You ask, not for my verdict, but for what I have observed, that you may judge for yourself. If I say, "I gave your applicant a '76'" and walk off, you have no observations but a perfunctory statement of my conclusions. Perhaps you are learning to drive, to paint, or to play the piano. You look to your mentors for guidance, not for their judgments. You learn to do these things because your primary interest and responsibility is with your own life and future, not just with outranking someone. If an instructor gives you a "C" every time you move, meaning commonness or mediocrity to almost everyone, you are not only not helped but you are discouraged, perhaps disgusted. Students submit to this disturbing process of labeling for years. That they survive is a tribute to their stamina, not to the teaching.

—Max Marshall 25-26

In the 1950s, Max Marshall a microbiologist from the Medical Center at the University of California, offered these words after a several year experiment in his microbiology courses. The no-grading practice spread to other science de-

partments. Over several years Marshall and his colleagues taught without giving grades in any of their courses. Marshall came to the above conclusions. Grades are “autocratic” and “despotic judgments.” They “introduce a serious pedagogic fault.” Students survive them out of stamina, not because of good teaching. I felt these same conclusions in my own teaching, and in part, they were the source of some of my returning to graduate school after a few years of teaching at the community college level.

In this chapter, I offer a narrative of my coming to labor-based grading contracts. I will not explain what contracts are in detail, but offer a kind of literature review that dramatizes my own problems and questions with the literature, brought on by my teaching. You might think of this chapter as a kind of background on the literature concerning grading that informed by problem-posing of grading and movement from hybrid grading contracts to labor-based ones. In another way, one can read this chapter as a fleshing out of Marshall’s conclusions about grading through a look at the literature on grading contracts that led me to my current practice.

AIN’T MAKING DOCILE-STUDENTS NO MORE

My problematizing judgment and assessment in my classroom took years to articulate itself and as my last chapter and Marshall’s words suggest, it began with the problematic of grades. I wish I’d found Marshall’s discussion at the front end of my teaching, during the mid 1990s. His humane respect for his students still strikes me as revolutionary, even if he does not consider the ways race, class, gender, or disability might affect his conclusions, might reveal the unevenness in the pedagogic fault of grades. Then again, I wonder how diverse his 1950s University of California students were? Early on, I didn’t have even Marshall’s language to think about grading, and I certainly didn’t have the language of whiteness or Foucault’s docile-making theory to help me make sense of what was happening in my classrooms’ assessment ecologies. I just figured that thoughtful, self-conscious, rhetorically-minded pedagogy could not be white supremacist, that if I made sure to treat everyone equally in my evaluations and grading, I wouldn’t be racist. But I was wrong.

During my time teaching at the community college, before going back to grad school, I also saw more obvious problems with grading writing, problems that others were discussing at the time (Tchudi; Zak and Weaver; Allison et al.). Grading seemed to get in the way of the conversations I wanted to have with students about their writing. It was a barrier when I would tell a multilingual student or a working-class white student, for instance, that her writing was interesting and moving, yet the C-grade seemed to say something else. Grades kept

us from having a real exchange. Most just wanted to follow my orders, or took everything I said as an order, but I wanted them to talk through my words, think with me about their writing, and make their own decisions. Many of my students seemed resigned to low grades and following orders, and I felt it wasn't all their fault. It was in part my fault. Beyond these pedagogical problems, I found myself hating to read student writing, even though I loved talking to students about their writing.

Additionally, I could hear the resignation and past beatdowns in our opening discussion each quarter. To introduce ourselves on the first day, I always ask something like: who are you as a reader and writer? Most students characterize themselves as bad writers or simply say they don't see themselves as a writer. Yet in my experience, many students, particularly women, will identify themselves as consistent, even avid, readers. This always seems paradoxical to me, even though I know that one may love reading but not find an urge to write. However, I still find it a problematic (not a problem) to enjoy reading words, yet not enjoy using them in other ways to articulate or discover what you think or feel, or to not see your own use of words in your life as thinking or writing, as legitimate languaging. This somewhat paradoxical stance, readers who don't see themselves as writers, affected how my past students read my comments when a grade was attached. No matter how encouraging I tried to be, the grade confirmed their sense of themselves as bad or substandard writers. While I wouldn't have called my classrooms docile-making spaces back then, they were.

I'm reminded of Kenneth Burke's famous definition of humanity. I'll adjust his references to be more gender inclusive:

[Humans are]

The symbol-using (symbol-making, symbol-misusing) animal

Inventor of the negative (or moralized by the negative)

Separated from [their] natural condition by instruments of [their] own making

Goaded by the spirit of hierarchy (or moved by the sense of order)

And rotten with perfection. (16)

If we accept Burke's definition of humanity, a large part of being human is our languaging, our ways with words, our symbol-using-making-misusing practices. These practices create the problematic reader/non-writer stance in my students through their very modes of expression. They say, "I am NOT a writer, but I am a reader," yet there they are languaging, making and using symbols to define themselves, separated—no, alienated—from their material conditions as language users by language itself, separated by their own expression of non-writ-

er status that they use to define themselves. How does this occur? Perhaps it is in the goading of the spirit of hierarchy enacted and symbolized through grades, in the yearning for the highest grade, in the disappointment of getting a lower mark. Why do grades have such power over us? Why are we moved by their sense of order? They suggest our imperfection, our un-success, our proximity to failure, and because we live in a society that functions from binaries, they offer us an indication, as imperfect as grades are, of how far we are from our own perfection, the ultimate motive for students.

So with this vague sense of something being wrong with my assessment practices, I went back to graduate school. I went initially to study rhetoric and racism with Victor Villanueva, but I ended up also studying writing assessment and thinking a lot about how to apply the ideas circulating in the literature on program assessment and large scale writing assessment to the writing classroom. I'd been a writing teacher for several years before returning to get my doctorate at Washington State University. I'd taught for a few years at Chemeketa Community College in Salem, Oregon, and before that I was a teaching fellow at Oregon State University, where I earned my M.A. in Rhetoric and Composition and was a Teaching Assistant. It was at OSU that I discovered my love for teaching, for helping others engage with words and create and communicate their ideas. I love pretty much everything about the writing classroom, except grading, of course. I love that it is a rare opportunity in which we pay attention to our words and the words of others. I love that the writing classroom is one of the only places we get to pay attention to how we cultivate meaning from and with the words and symbols that pass in front of us every day. When things are going well, the writing classroom is an open, engaging, and inviting place where we look at texts and artifacts and consider their ethical significance in the world and to ourselves. We sit in the presence of words. We slow our lives and minds down. We look closely. We listen carefully. We behold. We question. We ask. We sip meaning and significance from a deep well of language and culture. We care for those around us and their words. And from all this, we pose problems about our existential situations in our world of words and symbols.

But of course, my biggest problems in my teaching had always been grading writing. I didn't like doing it, and I've never met any writing teacher who liked doing it. I don't know many students who like grades either, which isn't to say that there aren't a lot of students who don't want them. My students back then at OSU, Chemeketa, and WSU didn't like getting grades either, at least not the grades I gave them on their writing. It ruined my relationship with their writing and even with them. In most cases, my grades ruined their relationship with their own writing. I found myself grudgingly reading every stack of papers because I knew that somewhere in that stack my mood and attitude would change.

I'd sour. I would have to justify each grade and that meant my stance as a reader would have to be one of looking for reasons why I didn't think the paper deserved a higher grade, or at best why I thought the paper warranted the grade I gave it—and I almost always knew the overall grade by the time I got halfway through each paper, which then turned my attention as a reader toward the markers in the paper that justified that grade, a self-fulfilling judgment practice. In short, I sensed that grading kept me from enjoying my students' writing. And as Peter Elbow argues convincingly ("Ranking" 200), I intuitively knew that I needed to like my students' writing, if not for their benefit, then my own.

Sure, I used carefully crafted and recrafted rubrics. I shared and worked through the rubrics with my students. I let them decide on their topics and kinds of writing they wanted to engage in. I demanded revision after peer workshops and my own feedback on their drafts before I put grades on papers. I tried to give them as much choice as possible, as much opportunity to know my expectations, but I still had to grade based on my expectations, even when those expectations were collaboratively crafted. The discourse of judgment and grading was based on a standard, one I translated in my judgments from our rubric. It all felt unfair to my students, not my expectations, but using them to determine progress and proficiency, course grades and GPAs, after we'd done so much collaborative work. And for my working-class, multilingual, and students of color, who often entered my courses with Englishes other than the dominant one I used as a standard, they were at a double disadvantage. They didn't have first-hand access to the standard. It is like playing a game in which you don't know all the rules until someone catches you breaking one, and you must then move back three spaces. It was a game I remember playing and somehow winning by a few lucky rolls of the dice. I was the anomaly, the fluke. It happens, but rarely. I won the game despite the rules that clearly were stacked against me. My students knew this game too. They knew the odds. I was now making them play it, telling them, if I can get lucky, so can you.

It was in Bill Condon's writing assessment course where I made my first steps toward contract grading in developing the article, "Community-Based Assessment Pedagogy." The course would literally change what I did and how I did everything from that point on, from my teaching to my research. Up to that point, I was studying rhetoric and racism. That was going to be my dissertation. But Bill's class showed me that racism is judgment, that the classroom is a site that reproduces racism and white language supremacy, that how judgments in such spaces are made have just as much to do with larger, structural forces as they do with an individual's idiosyncratic reading of a text. From the first class session, I was hooked. I saw connections to everything I did in the classroom and in my daily encounters with words. I began to see how racism is reproduced

through the ways we judge language that seem to be about everything else but race. We are always judging, always assessing, from racialized bodies and discourses. If language travels with people, if it is social and historical in nature, then it is shaped by the ways groups of people who use it are shaped. Language is racialized. We language through our racialized *habitus*.

On the first day of the semester, Bill walked into class and quickly handed out a page copied from an academic journal. He said we were going to engage in a reading exercise. He wanted us to count the number of e's on the page. Because we were all good doctoral students, we dutifully counted. A few minutes later, Bill asked us to go up to the board and write our answers. How many e's were there on the page? We all went up at the same time and wrote a different number. There were as many answers as there were students in the class. How could this be? What did it mean? I was shocked. Had I messed up my counting? Didn't we all have the same page? Yes, Bill confirmed that we did. He said if we, a group of smart doctoral students in an English Department, couldn't agree on the number of e's on a single piece of paper, a simple act of assessment, how could we possibly expect to agree on much more sophisticated and complex dimensions of the same page, like effectiveness or what it says, or how proficient it is? This was my entry into writing assessment. And through that class, I came to community-based assessment pedagogy, the practice and the article.

In the article, I argue for a cycle of rubric building, using, and reflection activities that lead to the use of a rubric in peer assessment activities, which culminate in a richer, more textured set of categorical assessments of students' writing without using grades on anything. At the end of the course, instead of the teacher evaluating the student's portfolio of work, the student collects several peer evaluations using our rubric along with their own self-evaluations. The student and I would sit down together and read all of them, then I'd read my own prepared evaluation of the portfolio, none of which would have grades on them, instead we had a simple categorical decision to make for each rubric dimension: does this student's portfolio meet, not meet, or exceed the expectations of our portfolio rubric along each dimension and how does it do that? These judgments and their reasonings were recorded in each assessment document.

In final conferences, we looked at the landscape of evaluations—peers, self, mine—and came to what I called “hard agreements” about what overall assessments we could make given what folks said about the writer's portfolio and the evidence they'd provided for those assessments. The final course grade was determined by deciding how many of the rubric dimensions met, exceeded, or didn't meet basic expectations along all portfolio rubric dimensions, which we'd also collaboratively created, discussed, revised, and reflected upon over the semester. If we could argue that more than half of the rubric dimensions

were exceeding expectations, then the student would receive an A, if most were meeting expectations, then a B. If there were a few not meeting, then a C. If more than half were not meeting expectations, then a D. If the writer's portfolio couldn't be argued that it met any of the class' expectations along any of the rubric dimensions, the student got an "F." This course grading system was still a quality-based grading system—that is, the final course grade depended on a discussion of the judgments of quality, albeit by several colleagues, the student-writer, and me.

In my estimation, while it didn't address the power differential between my students and myself in our assessments of portfolios (mine was always going to hold more weight in those final discussions), nor did the system account well for the ways some white and Asian female students didn't seem to resist as much the negative assessments of their writing by others, as white male students generally did, community-based assessment pedagogy was still a fairer system than conventional grading systems for three reasons. First, it offered students some added control and agency in what constitutes quality through rubric generation and reflection activities, as well as those final assessments. This leads to added agency in those discussions in the final conferences, where students get more voice in what their final grade ends up being. They get to argue, in a sense, or defend, their preferred grade, based on evidence. More importantly, they have a real and materially present voice at the table when the teacher determines course grades. Along with this, that course grade is also produced more collaboratively, not by one judge (the teacher) alone. While later I would see the intersectional, gendered, and racialized problems with various students trying to defend or argue for an assessment or a grade to an authority figure who is steeped in a masculine, white *habitus* of the university, at this early stage, this was still better than conventional power dynamics in teacher-only grading systems.

Second, community-based assessment pedagogy reduced the number of grade distinctions possible. The assessment literature is clear on this point, and it's more nuanced than what it seems at first glance. The fewer distinctions judges make, the more reliable those decisions are—that is, the more agreement there will be among various readers. Furthermore, my system asked students to think differently about these kinds of judgments. Because we weren't adding or averaging points or percentages, we could think more conceptually, closer to the way people often judge language organically in other places outside of classrooms. When we judge texts, arguments, advertisements, movies, or other discourses in the world, we rarely place them on linear, numerical scales. Instead, our judgments are more conceptual. We may not have explicit rubrics or lists of expectations, but we do work from expectations, known examples, and salient features that we read as quality or something important to make a decision from.

We say, “that commercial is so retro. I love it,” or “that song is interesting in a backwards way, like the old Beatles tunes,” or “I found the speech by the President baseless and unconvincing because he didn’t give any reasons for any of his policy decisions beyond ‘because I said so.’”

I chose to resist linear models and judgments that amounted to points because I felt that our decisions were more outcomes based. That is, we had a set of agreed-upon outcomes, our rubric, so we could simply make a categorical decision in each case, each dimension. Did the portfolio meet expectations in this area? We mostly just needed to know that. Then in some cases, we needed to know if the writer exceeded expectations. Was the writing doing something demonstrably more than most everyone else in the class?

Yes, I’m mixing assessment paradigms. The initial categorical decision could be seen as a judgment much like criterion referenced tests offer. Criterion-referenced tests purport to show how well a person does compared to a particular skill, outcome, or construct already identified, such as the expectations listed on our rubric (the criteria). Meanwhile, I was urging students to make the higher decision, whether a portfolio exceeded expectations along any given rubric dimension, using a norm referenced model of judgment. Norm-referenced tests purport to measure test takers relative to other test takers. In other words, one is exceeding expectations based on the reader’s sense of what other writers in the class are doing (or not able to do).

I still think this criterion-to-normed reference judgments is a good way to help students, even faculty, make such decisions, depending on what those decisions are used for. But there is another way to see how these judgments work at the micro-level, the level of the individual reader-judge—categorization theory—and this theory likely was floating in my head at the time because we used it explicitly in the writing program at Washington State University to judge placement essays and junior rising portfolios, processes I was a part of each year. Categorization theory explains how our minds make decisions and judgments like these, and gives insight into what I was trying to lead my students through. In the context of holistic scoring, Richard Haswell, a former WPA at WSU, explains the three main types of categorization that psychologists say we make judgments by: classical, exemplar, and prototype. Classical categorization says that we judge by looking for the “non-accidental properties of a new instance and matching them with the unique set of properties that define the correct category” (245). Each category then has a set of known, ideal features and a reader attempts to match the features seen or heard in the present text with these known, ideal features. Our collaborative rubric attempted to help us articulate such categories and features, only from inductive methods, which means our process of rubric creation shared another kind of categorization, exemplar.

Exemplar categorization says that people “categorize by comparing a new instance with intact memories (‘exemplars’) of similar instances” (Haswell 247). Thus we read a text looking to see how closely it fits to significant features of a most representative or best example of the category. Our rubric-building process assumed a kind of reverse method for producing the categories and descriptive features. It began with actual known exemplars, texts we brought to class, ones we wanted to imitate as writers. Through a series of activities, we derived dimensions of writing from them and those dimensions’ features, which we then used to help us make judgments on each other’s drafts and portfolios in the assessment documents. This process was influenced by Bob Broad’s dynamic criteria mapping process (see Chapter 5 of Broad). In use, the actual exemplars that our rubric was based on go away, and what remains is a bricolage of exemplar-inspired dimensions that we use to remind us of our priorities when assessing.

It is prototype categorization that I asked students to engage in during our final assessment activities on those final portfolios. Prototype categorization says that we “judg[e] how similar the yet-to-be-categorized instance is to abstract schemas [we] have of the best example or most representative member (prototype) of possible categories” (Haswell 246). The prototype for any category is, then, an “idealized construction,” a “convenient grammatical fiction” (Haswell 246; Rosch 40). So we don’t actually use a real example when we judge instances, instead we use convenient fictions, prototypes in our heads that are cobbled together from various examples. Our rubric, then, was heuristical in nature, reminding us of key categories and features of the prototypes in our heads that we all agreed upon to use when we built the rubric, which was a kind of classical set of categories. What complicates this is that what any individual judge can see and discuss will vary a bit, even though we are sharing the same rubric and coming to agreements about what categories mean. But since we always work from our own mental, idealized constructions of things, this paradigm of judgment seems to explain best what we were trying to do and how it would inevitably be uneven in application—there would be necessary disagreement. Each reader was reading for and explaining how close a given portfolio was to the construction of an idealized portfolio that existed in that reader’s head, no matter the rubric we’d collaboratively made.

While we were using prototypes in our heads that we articulated on our rubric and agreed upon as a class, each student’s (and my) expectations were still idiosyncratic by their nature. Remember, our judgments have patterns in groups because they are structurally determined, which our *habitus* help account for. Those *habitus* are a part of our language systems since our language systems are a part of our racialized and social histories in society, schools, and academic disciplines. In short, readers are constrained and pressured in particular directions to see and value particular features of texts no matter the prototypes we hold in

front of us for judgment purposes. This means readers who share *habitus* will be determined to judge the same text in similar ways, but not exactly the same ways. We have agency, just not full freedom to see things in any old way. Without examining our rubric and prototypes in structural and historical ways, it will be hard to avoid reproducing blindly a white racial *habitus* in any writing classroom or rubric, thus difficult to not privilege those who already embody such a *habitus* in that classroom through our judgments of writing. This is the exact dynamic I was trying to avoid in my assessment systems. Community-based assessment pedagogy seemed to easily reproduce unfair, racialized dynamics, and this made it another structural way that writing assessment reproduced white language supremacy, despite my better intentions and my own racialized history of subordination in school. Thus as long as I had to produce grades, I couldn't account for individual variance and difference in judging language.

To summarize this second advantage, community-based assessment pedagogy asked students to think of the judgments of meets, exceeds, or not meets expectations as conceptual buckets, or categories, as discrete partitions. The writer and I decided which category each portfolio fit into after hearing all the individual assessments of the portfolio. The system resisted harmful linear, hierarchical scales, determined only by a teacher, yet there were contradictions. Our categorical method resulted in students still being placed into a hierarchical system of course grades from A-F, a system of partitions that still functioned as the art of rank, only it allowed students to control some of the movements in the enclosures created. The categories themselves were partitions in the enclosures of the rubric and portfolio. I just got my students to manage those enclosures with me. Many of these contradictions occurred because the university required grades. And of course, there are always contradictions in any assessment ecology situated in other ecologies that demand things like movement between enclosures that are already created.

The third advantage, and unexpectedly, of community-based assessment pedagogy was the way it began to change the culture of my classrooms and interactions with students. I'm sure some students saw the ecology as fairer, but many still resisted. But I began to enjoy being in the assessment ecology, just like I had remembered loving the classroom in my early days of teaching when I was so intoxicated by this new and exciting space, these new and exciting practices, that I didn't have time for the contradictions. And because I wanted the assessment of writing in my classes to be a mutual practice of students and me working through problems of judgment in language together, posing problems together, learning together, having conversations about language and how we judge it differently, I saw promise in this more student-centered assessment system, one that had fewer distinctions in assessments at the end of the course. I was getting close, but something was missing.

What was difficult to consider in this first iteration of my own assessment practices was understanding the nature of judgment and seeing other possibilities for producing course grades. I was fortunate enough to get feedback from Peter Elbow on the article. He was one of the reviewers, and he signed his review and provided his phone number, suggesting I call. He had an idea for me that he felt might help me. Of course, I called. I was deeply flattered that such an academic rock star would have much to say to me. I had cut my teeth on *Writing Without Teachers*, which had been suggested to me by Chris Anderson, my first mentor and the Director of Writing at Oregon State a decade earlier. So when Peter Elbow suggested I call him to talk about my ideas in the article, I called. He gave me his grading contract, a version of the one he and Jane Danielewicz published later (Danielewicz and Elbow), but Peter had been using and working on contracts and alternative systems of judgment in writing classrooms for at least a decade before (Elbow, “Grading Writing,” “Taking Time Out,” and “Ranking”). I tried the grading contract the very next semester and never turned back. Peter was right. It solved many problems around grading that I would still be having, and opened up many other possibilities. That first semester was in 2004. I’ve not graded a paper since the spring of that year, and my students and I are better for it.

FIRST TRY, HYBRID CONTRACTS

The grading contract I inherited from Peter Elbow judged two kinds of evidence from students’ learning in the classroom to determine some course grades (those above the default B grade), which Danielewicz and Elbow refer to as a “hybrid grading contract” (250). They explain that course grades up to a B are contracted by doing particular things, by “conscientious effort and participation” (246). In early versions of my contract, this part of the method was described as “behaviors,” much like Danielewicz and Elbow’s contract describes things. I promised students: Behave in a particular way described in the contract, do the work asked in the spirit it is assigned, and you are guaranteed a “B” course grade, no matter what I think of your writing. I took this language from the contract Peter shared with me. But A grades required different evidence to be judged.

While most course grades were determined by effort, participation, or behaviors, grades above a B required a judgment of writing quality, usually on a portfolio at the end of the course. Again, this feature was taken from Elbow’s contract given to me. In Danielewicz and Elbow’s contracts, this higher course grade is determined by the teacher’s judgment of those final portfolios that rose to the top upon final reading of them. They explain this in clear terms: “We need only examine the remaining final portfolios that are particularly strong to decide which students get which grade higher than B. Often there are not so many”

(253). They offer ways to make such judgments explicit, which many students will care about and want. They explain that a course grade of B “should be available to *every* student—that is, not dependent on skill or prior training”—while A grades require “specific criteria or features of ‘good writing’ that not every student could attain” (258; emphasis in original). In other words, A grades require judgments of writing quality that are made explicit to students, but are decisions based on a teacher’s judgments of the literacy performances of those students.

Hybrid contracts’ basic premise is clear: your assessment ecology will keep the most exemplary grades away from some students, regardless of their desires for those grades or the amount of work they are willing to put into their writing in order to get those grades. More often than not, these students will be students of color, working-class, and multilingual students—students with *habitus* other than the dominant white racial *habitus* embodied in the teacher, rubric, standards, and course outcomes. Thus, grades of A in Danielewicz and Elbow’s hybrid contract system are determined by a teacher’s judgment of the *habitus* that are embodied in students’ languaging. It didn’t take long for me to feel uncomfortable about this problem in hybrid contracts. And it amounts to hybrid contracts being subtly racist and white supremacist.

As community-based assessment pedagogy argues, I’ve never liked the idea of only the teacher grading student writing, and I’ve always been suspicious of any one judge deciding what is exemplary literacy performances, even if that judge is a teacher. One lesson anyone can see from academic and civic work is that fairness is a construction of the community involved, the people most affected by decisions. If you get to have a hand in the decision that affects you, no matter what it is or what it’s about, you likely will feel that the consequences of that decision are fairer to you. Grades in classrooms work the same way. So my first iteration of hybrid grading contracts asked for several student colleagues, the writer, and myself to provide written assessments of the final portfolio, just like my community-based assessment pedagogy. The writer and I then got together in a final conference to read through and discuss this landscape of judgment. Up to this point, I thought the key to fairness was in collaboratively building rubrics and deciding final grades together with each student. I thought constructing fairness was about participation and method. But these things are also tied to the nature of the decisions, those decisions’ consequences after the class, and the shifting, intersectional, and uneven social structures and power relations that construct my students and me.

With the hybrid contract placed inside my existing community-based assessment pedagogy, I found that the student and I didn’t need to discuss or make all those distinctions at the end of the semester since all grades up to a B were already determined, and all that was left to determine was the quality of writing by those few students who may qualify for A-grades. However, in my system, the A-grades

were collaboratively determined by considering carefully and with the student all the assessment documents on their final portfolio. So instead of five grade distinctions to make, students and I now only had one. It was more elegant and simpler. There was more agreement and fairness, since students helped figure the grade out.

After a few years of using hybrid grading contracts in this way, and teaching in racially and culturally diverse classrooms, the racism and white supremacist tendencies with this hybrid contract became clearer to me. It wasn't just a hybrid system. It was a fundamentally unfair system to many students of color, multilingual students, and students who came to my college classrooms with Englishes that were distant from my own and the ones I was asked to promote, that is, dominant academic discourses (or DADs).¹⁰ As Danielewicz and Elbow say, some students will not be able to attain an excellent or superior judgment in these dominant discourses because their literacy practices are just different—they didn't grow up in white, middle-class households—and ten or fifteen weeks of instruction is just not enough time to change these linguistic realities. So many of my students of color at Fresno State didn't have a chance to get an A, even if they wanted it and were willing to work very hard for it. This didn't sit well with me. It still meant that the same students who benefit from conventional grading systems that use a white, middle-class standard discourse are privileged and placed at the top of the grading pyramid and rewarded for less labor because their white racial *habitus* is the center of all DADs. My DAD was dictating who got the highest, most coveted grades in my classrooms. Meanwhile, my students of color and multilingual students often worked harder but achieved a lower course grade.

At an experiential level, this hybrid contract was perhaps even more unfair than conventional grading. What I was telling my students from the first day is that our contract allows us to let go of grades, to focus our attention on learning and the feedback we get on our writing, and making our own informed decisions in our writing—not follow orders. My motto was, and has always been: Good writers make careful decisions; they don't follow orders. I would tell them that you don't have to worry about me not liking your writing. Listen to what I say, try to do all you can, but make your own decisions and have good reasons for why you want to do what you do in drafts. I was trying to give them a right to their own languages, but the grading system still didn't fully allow this. If a student took seriously my words, if they enacted their own right to their own language, then they only had a right to a B-grade.

Now, at least students who actually took my words to heart couldn't fail the class. Yet at the end of the semester, when it really counted, particularly for those

10 I often begin talking with my students about dominant standards of English by referring to them as “DAD”—our DAD is our dominant academic discourse—but DADs are really the dominant white *habitus* that are embodied in language standards and norms.

who wanted an A-grade, they really did have to care about their colleagues' and my judgments of their writing—they had to follow orders—which always meant imitating a dominant white racial *habitus*. There was a contradiction, a racist problematic: All students had a right to their own languages, just not a right to an A, not a right to have their languages valued most highly, unless that language matched a white racial *habitus*. So hybrid grading contracts still participated in white language supremacy.

This core assumption in hybrid grading contracts maintains the hegemonic racial and linguistic hierarchies we see today. It does nothing to change them. It is white supremacist, since it reserves the highest categories of grades in a system that rewards students with such grades, to those demonstrating a dominant white racial *habitus*, yet it seemed to suggest that the assessment ecology was helping non-white students achieve by not failing them for their efforts, even when those efforts did not amount to mimicking the standard. To use the metaphor of the early twentieth century, racist, separatist, and eugenicist, Lothrop Stoddard, in *The Rising Tide of Color*, hybrid grading contracts maintain the white center by managing the “inner dikes” that protect them. The white center is the A-grade partition, while the inner dike is the B-grade territory, open to all in the contract, but carefully separated from the A-grade by maintaining the dominant white racial linguistic standard. This means the “outer dikes” are the less-preferred, lower-grade-lands.

I doubt many would characterize the differential judging in hybrid contracts as white supremacist in the ways Stoddard clearly thinks of geopolitics and their connection to the preservation of the white race and white lands. And yet, by default, hybrid contracts work to produce the same inner and outer dike system, controlling movement through both. The grading philosophy, to me, sounds like this: Let the Brown and Black folks have their Bs and Cs. Let them pass writing courses. Meanwhile, the As, the real opportunities and high GPAs, will be reserved for those who can embody a dominant white racial *habitus*, which means white students get overrepresented in the A category. The world and its educational opportunities remains guarded from the rising tide of color in our schools through the system of inner and outer grade dikes.

In my own classrooms, what complicated the use of hybrid contracts further was that my good intentions to have students involved in the judgment of writing at the end of the term meant that individual student judgments would also reflect the valuing of a dominant white racial *habitus* over others. I could and did get my students—even students of color—to voice their own oppression. This is how we internally colonized our students. We get them to enforce the partitions and dikes of grading with us.

Ultimately, I couldn't see how this hybrid contract *uncompromisingly* valued students' right to their own languages through the *judgments of writing*. What it did was allow for some affordances around language diversity, up to the B dike. A student who wrote in Black English or Spanglish could pass the class with a B, but couldn't realistically get an A. It maintained the racial hierarchies that privileged dominant white racial dispositions to language. As good as this hybrid contract was in helping many students do better in the class than they likely would have with traditional grading, it ultimately gave them the wrong lesson if they used their non-dominant and non-white discourses. The lesson was that anything other than the white DAD of the classroom was less than, was sub-standard, and was not excellent writing. This meant that those of us who didn't grow up with a white DAD, likely couldn't get the A—my former, student-self could not get an A, even in my own class. As a writing assessment ecology, it didn't do the kind of social justice work I envisioned for my classrooms. Social justice writing assessment projects make people's lives better, give them equitable opportunities, expand boundaries, question the authoritative and hegemonic, not simply give some students a few more opportunities, or a second-class status, even though that status is better than the third-class one they are accustomed to. I wanted my contract to do more. I wanted it to be uncompromising when it came to the racist and white supremacist system my students had to learn in.

MOVIN' TO QUANTITY OVER QUALITY

During this time, I began doing more research on grading and learning contracts. Most published work up to the early 2000s discussed learning contracts, not grading contracts, and came from education circles (e.g. Berte; Knowles; Anderson, Boud, and Sampson). Researchers tend to discuss learning contracts in terms of K-12, adult education, or nursing contexts. Often, the articles and books avoid any detailed discussion of how course grades are exactly determined. This makes sense. Learning contracts are designed to help learners achieve particular learning outcomes, not grades. They are more concerned with helping students master particular tasks or outcomes, like setting a broken leg or diagnosing a patient.

For instance, Margaret McAllister discusses learning contracts for nursing students in Australia. Each individual student's learning contract consists of "learning objectives," "resources and strategies for learning," "evidence of accomplishment," and "criteria for evaluation" (200). While the last two areas of McAllister's learning contract offers ways to understand how a student will be evaluated, she says very little about exactly how the evidence of accomplishment marshalled by the student will actually be used in an evaluation process to then calculate a course grade.

In using learning contracts in self-directed, second language learning classrooms in Singapore and Canada, Hedy McGarrell offers similar ways to structure contracts in English language learning classrooms. The main elements of McGarrell's contracts address:

1. "What will the learner learn?" (497);
2. "How will the learner accomplish an identified short-term objective?" (498);
3. "When does the learner expect to achieve a given objective?" (498);
4. "How will the learner demonstrate what he/she learned?" (498); and
5. "[H]ow will the learner's demonstration of achievement be evaluated and by whom?" (498).

While no specifics are given toward calculating grades with learning contracts, McGarrell does push for students to be a part of the evaluation process. She explains, "principles of self-directed learning require learners to evaluate not only the achievements of their learning objectives but also the quality of their learning experience" (498). She goes on to say that some teachers may find it helpful to evaluate the student's performance, but should avoid "instructor-centered techniques." Her suggestions amount to the teacher "remain[ing] a facilitator, rather than a judge, of the learning process" (499). McGarrell promotes a kind of assessment ecology that I was (and still am) looking for, one that gives students more control over the ecology and is flexible for a variety of learning paces and styles. But there is a difference between allowing students control over the evaluation of the evidence of their learning and allowing students to control how their course grades are calculated from judgments upon that evidence. As Stephen Tchudi reminds us, there are differences between response, assessment, evaluation, and grading (xv). And this is why, I'm sure, McGarrell pushes for an avoidance of instructor-centered evaluation, but it's unclear if that includes grading. Furthermore, it isn't clear at all how grades are calculated in her classroom. She remains silent on any options. Of course, learning contracts do not contract for grades. They contract for learning, so there is no reason she should explain this.

The value in learning contracts, I think, is clear. They focus on negotiated learning processes and outcomes or goals for individual projects and are individualized to each student. In *Learning Contracts: A Practical Guide*, Anderson, Boud, and Sampson define a learning contract as

a document used to assist in the planning of a learning project
... a written agreement negotiated between a learner and a
teacher, lecturer or staff adviser that a particular activity will
be undertaken in order to achieve a specific learning goal or

goals . . . a learning contract is essentially a “process plan.” It is a means of designing a learning activity with the focus on the learner rather than the subject or the teacher. For this reason learning contracts are particularly suitable for structuring assignments and projects which are largely self-directed, for use in courses in which participants come from a diversity of backgrounds and in tailoring learning to individual needs and interests. (2-3)

Thus learning contracts promote student responsibility, agency, and control in mostly self-governed, learning processes, with an attention to what that process is meant to achieve or produce, and how that learning is evaluated for success or completion. They promote metacognition through the ways the learner must help develop the contract with the teacher, articulating what goals they wish to achieve, what evidence of success looks like, and in some cases, even assessing their own completion of the contract. But most of the literature on learning contracts do not give students any explicit control or responsibility for their course grades, which for students in most contemporary college settings is intimately linked to their learning.

While educators might agree that learning and grades are separate, we cannot just tell students this and expect that they will understand how they are separate, or be willing to act on such an understanding. Why do we think our students should trust us when we say their grade and their learning are separate while everyone and the system itself up to that point in school has acted in contrary ways? In fact, acting on this information may harm them in our classrooms or others. This makes learning contracts different than grading contracts in this key purpose. Learning contracts make agreements about learning processes, while grading contracts make agreements about grades.

The relatively few articles that were published on grading contracts in college classrooms come mostly in the 1970s, '90s, and early 2000s, sprinkled here and there. There have been some articles on grading contracts as modified, criterion-referenced systems in higher education (Hassencahl), some of which look at students' favorable reactions to them in teacher education courses (Taylor; Brubaker), college nutrition and food management courses (Blankenship), and agriculture education courses (Newcomb), while others consider their success in accounting courses (Zarzeski), larger psychology courses (Kirschenbaum and Riechmann), as well as speech communication (Stelzner; Wolvin and Wolvin), and nursing courses (Kruse and Barger; Crancer, Maury-Hess, and Dunn). Some have discussed grading contracts' uses to enhance critical pedagogies (Shor, *Empowering Education* 159; “Critical Pedagogy”) and postmodern peda-

gogies (Hiller and Hietapelto), or simply as a more efficient method for calculating grades that helps students write more (Leahy). Most in one way or another, make a point to discuss the democratizing elements of grading contracts, giving students more control in courses and power over their grades (Danielewicz and Elbow; Inoue, *Antiracist*), as well as authority in the classroom (Burbaker). There also have been studies about grading contract systems in college business courses that show a statistically significant increase in motivation and effort (Polczynski and Shirland), while more recent studies have shown mixed results from first-year writing students (Spidell and Thelin; Inman and Powell) and statistically significant differences in performances and attitudes about the contract in students of color in a writing program (Inoue, “Grading Contracts”). All of this scholarship on grading contracts, however, uses hybrid contracts.¹¹ I’ll discuss in more detail Spidell and Thelin’s and Inman and Powell’s studies together in Chapter 7, where I discuss two criticisms of hybrid grading contracts that could be leveled against labor-based ones, so I’ll avoid my critique of them here.

The central assessment issue at stake in most grading contract discussions is about the nature of judgment that produces any given grade on a written performance. Most discussions assume a binary of quantity vs. quality (Bauman 165; Shiffman 67-68; Reichert 60; Danielewicz and Elbow 250). The two competing philosophical assumptions about the practice of grading and its relation to students’ learning to write might be stated like this:

Learning to write or improving students’ writing requires teachers to *judge writing quality*, thus course grades should be calculated by those judgments of quality. To be a fair assessment ecology, consistent judgments of quality are central.

versus

Learning to write or improving students’ writing requires students to *produce a certain quantity* of writing, thus course grades should be calculated by the quantity of writing a student produces. To be a fair assessment ecology, consistent judgments of quantity are central.

As mentioned already, Danielewicz and Elbow describe their contract as a hybrid model, one that uses both kinds of judgment to produce various kinds of grades. Judgments of quantity are used to calculate course grades up to a B,

11 The exception, as far as I can tell, is my own scholarship. While not all the grading contracts in “Grading Contracts” were purely labor-based, most were, and my own in *Antiracist* was. But of course, I didn’t have this scholarship available to me when I was looking for a better contract system.

while judgments of quality are used for those above a B. I found in my own classrooms that hybrid contracts like Danielewicz and Elbow's were unfair and white supremacist because they changed the rules of the game in the last inning, when the score was tied and there was a runner on third.

Savvy students figured this out quickly, and rightfully ignored my advice during the semester and instead tried to give me what I wanted, since clearly I was the most important judge in the room. I would be the one who needed the most convincing when final conferences came. They were negotiating with me. My students of color, those who usually found themselves at the lower end of the quality scale, were hurt more often in this schizophrenic, hybrid system of grading.

After just a few years at Fresno State, I decided to try to figure out how I might use quantity judgments only in all contract grade distinctions. There is some literature that supports this kind of assessment ecology, which I began to look at more carefully. In 1976, John V. Knapp offered a version of a quantity system to produce course grades, a grading system in which students receive a simple binary judgment on drafts, acceptable or unacceptable (Knapp 651). Once a draft is deemed acceptable, the student may move on to the next essay in the class. If they don't want to keep trying, they abandon that essay and move to a new essay. There are eight essays assigned, and over the course of the semester, students acquire as many acceptable drafts as possible, with three to four giving the student a C course grade, five to six a B, and seven to eight an A. The more writing a student accomplishes in Knapp's system, the higher their grade.

While Knapp's system avoids the teacher making too many judgments of quality on individual drafts—he only makes a single, binary one on any given piece of writing—it still uses judgments of quality by a single teacher. It is still a hybrid model. It avoids most ranking that occurs in grading and point systems, but ultimately maintains a hierarchical system of quality based on the teacher's idiosyncratic DAD, which categorizes students into acceptable and unacceptable. Any labor put forth by the student will not do. It must be labor that equates to an acceptable draft judged by the teacher. Knapp's system too easily works from an unquestioned, dominant white racial *habitus* that informs the dispositions to language that writing teachers always carry with them, no matter their intentions or who they are. The foundation of those simple binary judgments of quality that Knapp uses to make judgments of quantity is still white supremacist. His system, like the hybrid contract, protects the white center, the white literacy property, and creates unspoken inner dikes.

So I wondered: how do I avoid the white supremacy in judgments of quality on which Knapp's system still relies? Is it even possible to escape the hegemonic, white supremacist system of judgment in schools and colleges? Knapp's system

seems a fairer system, one that rewards those who do more work in the class. And those judgments of quality that he does make have a low threshold. In order to pass, a paper just has to meet minimum quality standards according to the teacher. And yet, everyone's labor is not equal in this system. A multilingual student working 10 hours on an essay may not meet the minimum standard for a passing essay, yet a white, middle-class student, who was raised in a monolingual English home, might spend five hours on the same essay, achieving a pass and moving on. Because the teacher's judgments of quality are what determines whether any draft is acceptable or not in the system, some students may still not be able to achieve a high grade, even if they desire to and are willing to work extra hard. To draw on John Rawls' idea of "justice as fairness," I wasn't convinced that a system that depended on my own judgments of minimum quality of drafts, which was by default comparisons to a dominant white racial *habitus*, gave all students a "fair equality of opportunity" (43).¹² I realized that in a hybrid contract, judgments of quality protect the white center of literacy as white property.

Beyond this problem, something else bothered me about Knapp's system. With my judgments of quality still regulating the assessment ecology, students would be too constrained, still likely to default to following my orders. In fact, the system is based on it. And I just didn't want them to do that, at least not by default. I wanted them to have good reasons to do what they did with their language and make their own decisions in writing. Perhaps unrealistically, I wanted them to be free to write, to write in a free and uninhibited place, to pursue languaging practices in ways they felt to be worth their time, which could mean learning dominant white racial habits of language. I knew that even my minimum expectations of quality, steeped in a self-conscious white racial *habitus*, would taint that ecology if used to grade, keeping many of my multilingual and students of color from more fully exercising their agency through languaging, and would likely offer more freedom to my monolingual, white, middle-class students. So I continued to search.

DISCOVERING THE PRIMACY OF LABOR

Through my research and reflection, I realized that what I value most in students is their working, their labor. This likely comes from my own poor, working-class

12 Rawls explains that "fair equality of opportunity" means that opportunities in a society or system are (1) formally open to all, and (2) "that all should have a fair chance to attain them" (43). I realize the critiques of Rawls' theory of social justice by scholars such as Iris Young who argue for more structural understandings of fairness, but I think taken on its own, this bit of Rawlsian theory holds up as structural.

background, as well as my own particular *habitus* in school. As a student, I knew I'd never be the smartest kid in the class. My grades seemed to prove that to me. How white people in and out of school talked down to me and treated me also suggested this to me. I can remember just showing up to neighborhood events, Halloween parties, block parties, and seeing the disgusted looks on the adults' faces by my mere presence. They were all working-class whites. I was a Brown kid, Japanese, to them Mexican. I was 13, 14. I lived in an explicitly racist world. The racism was very present to me. Most of the time, it wasn't even veiled. I felt this same tension in school. During my Freshman year of high school, I got an A in honors French and every other class I took, yet received a B (not a B+) in English, not honors English, regular English. How was this possible? What was I doing wrong? Apparently, nothing. It was me, my *habitus*. I knew this but didn't want to admit it, admit that my language and body were being judged together. This sense of me against the world helped cultivate a stubbornness in me. I wasn't gonna let those white folks get the best of me. They can't stop me from working hard. At least, I have my labor.

The summer before my Freshman year of high school, I figured I may not be rewarded with all A-grades in school, but I sure as hell could be the hardest working motherfucker in the room. And so, that is what I did. I wrote lists and schedules. I was diligent and careful in my planning of my time. I started things early and worked longer on everything. I did not—and to this day, still do not—procrastinate. Now, to have this as your philosophy for achievement and success in school and to not always be rewarded for it meant that I needed some other motivation beyond grades, more motivation than sticking it to the man with my As. My motivation was the work itself. I figured, shit, I'm spending most of my free time on school work and studying, I should try to enjoy the doing of it. I won't lie and tell you that I was an enlightened soul at 14. I wasn't. But looking back, I was cultivating a privileging of labor over quality from age 14 to graduate school, in part out of necessity. So it really is no wonder that by the time I was a writing teacher, I saw problems with the clear and present contradictions in my own classrooms between my students' labors and the grades I was giving their products of those labors. It was unfair. I knew this, felt it—had felt it for years! It just took several more years to figure out the problematic.

To put it bluntly, to produce course grades, I trust my judgments of the quantity of writing more than I do of the quality of writing. But I realize that mere quantity doesn't necessarily equate equally to time spent laboring over a draft. Some people are faster at producing a lot of text, some are slow typists, some pine over every sentence, some just throw every word in their head down on paper, piling up pages quickly. There is lots to unpack here: we aren't simply talking about quantity or quality, but my judgments of such things. What makes more

work better than so-called quality of work in a classroom where the goal is learning to write in English? Isn't it true that often less is more? How does quantity as a primary value in an assessment ecology address social inequalities in the nature of various students' languages or discourses? What's the relationship between focusing on quantity and the quality of literacy performances, or better yet, student future success? If using judgments of quality is white supremacist and racist, what makes using judgments of quantity of writing any fairer or antiracist? What about students who have other demands on their time, intersections of class and economics, intersections that surely played a role in my own background? Aren't there students who likely don't have to work and go to school at the same time? Won't they be just as privileged in a purely labor-based grading system where arguably time is the key factor for success as in typical quality-based systems of grading? Aren't those more time-privileged students also more likely to come from more economically well-off families, and aren't those families statistically more likely to be white families? These were, and still are, my concerns and questions. I don't have clear answers, even today. I do have some answers that have worked, but not fully. These questions continue to be a problematic.

So in my rethinking of grading contracts, I saw Knapp's binary judgments, which may never be avoided, as both an innovation and a problematic: judging quantity by judging minimum quality is better than having judgments of quality only, but this still ignores the work put into drafts. It ignores the labor. Nancy Reichart attempts to address this issue in her contract system, embracing both kinds of judgment. Her version of hybrid grading contracts articulates both quantity and quality criteria for students to meet in order to get the assigned grade (63). Unlike Danielewicz and Elbow's hybrid contract, which uses only quality judgments for grades above the default B grade, Reichart's uses quality judgments together for all grade distinctions. Her system is similar to Knapp's, only more explicit about what counts as minimum quality. Reichart's contract attempts to be clear about the fact that in order for a draft to be complete and count, even if a student is just trying to do the minimum, it must meet a particular set of quality standards. These standards though are still based on dominant white racial *habitus* and standards of "good writing" determined and judged by the teacher. How can these kinds of binary judgments be done in such a way that check my own version of a white racial *habitus* and discursive biases, while still using them because they're the main ones I have to judge language?

One early discussion that was particularly helpful in solving this quality vs quantity problem was Barrett John Mandel's "Teaching Without Judging," which was published in *College English* in 1973. Essentially, Mandel's system is one based on quantity of work completed by students. If a student wants a C in the class, they attend class regularly and turn in a complete C-level essay.

The designation “C-level” refers to the requirements of the essay, not Mandel’s judgment of quality. For Mandel, C-level work was a collaboratively written draft that students wrote and submitted. If an individual student wanted a B in the course, then they turned in B-level work. The range of B-level work Mandel listed were things such as “an intellectual journal covering the course readings and the class itself. Due twice: mid-term and end of the term,” or “close analysis of one play (if the analysis differs from that which evolve in class discussion). Due any time” (Mandel 629). So to get a higher grade, Mandel’s contract asked students simply to do more work in the class. What I took from Mandel’s system was the idea of a contract that used only judgments of quantity, not quality, to determine all course grades. The more work or labor a student does for the class, the higher their grade should be.

Now, one might think that Mandel, like Knapp and Reichart, still had some minimum threshold for quality of any writing that meets the requirements for A- or B- or C-level work. According to Mandel, this is not the case. He accepts whatever the student hands in, but of course, offers his feedback on it for improvement on the next assignment. He explains,

I will, theoretically, *accept* trash submitted for an “A”. But I *believe* that in a non-judgmental, unpunitive, encouraging context, students will want to work toward achieving self-styled and often very challenging goals. While nothing in the format of the course coerces a student to do anything which reason, energetic teaching, and the student’s native curiosity do not inspire, I, needless to say, constantly encourage self-discipline and self-respecting work. (629; emphasis in original)

Mandel calls this having “faith in students,” which he admits is easier for him than many other teachers (629). On occasion, he says, he gets the “rushed or careless junk. But [his] approach to teaching is geared to those who can and want to learn, no[t] to those who, for reasons they are entitled to, cannot avail themselves to the opportunities to learn” (629-30). In short, Mandel designed his assessment ecology around students who are willing to have faith in the ecology and in themselves to learn. He feels no need to keep his students in line, to force them to turn in writing of quality that he determines is acceptable, instead Mandel’s ecology assumes that students will want to form lines in the ways they know how, that they won’t need to be held accountable since they can be responsible on their own. They want to learn and want to work.

When I first read Mandel’s article, one published when I was only three years old, I couldn’t help but find a deep affinity to his ethical stand concerning his

teaching and students. It struck me then and now that my students are entitled to give me whatever they want—or rather, whatever they *can*. And I’m equally entitled to tell them what I think of it and how it may not be helping them toward their goals in the course. They may be ready to listen or not, but their learning and labor should be their choice. I cannot coerce them into learning or enlightenment. I agree with Mandel still. This is how we all should be able to work with students, to labor together. I realize that all teachers may not be able or willing to try “energetic teaching” that inspires students to do “self-disciplined and self-respecting work,” or that if we do attempt this kind of teaching, we could recognize such efforts by all students, or even that such energetic teaching will always lead to students responding in these ways. I also realize the problematic in notions of teaching that promote unquestioned ideas about students that just need freedom and encouragement to do what they want in order to really learn, yet classrooms require environments that encourage risks and learning through failure, doing what we can, not what others tell us to. While I think Mandel’s system is not overly-idealistic about his students, I do realize that to many, it seems like pie-in-the-sky, dreamy pedagogy.

I also know that Mandel’s system sounds a bit like hippie-free-love-anti-teaching, and this may discredit it in some teachers’ eyes. But I ask why? What is so wrong with a “non-judgmental, unpunitive, encouraging” classroom assessment ecologies? Who says that judgmental, punitive, and discouraging assessment ecologies work better? I know, I’m creating a false binary, but I’m also trying to reveal what our real assumptions are about the kinds of judgments we make on student writing, how we use them, and why we think those uses are the best ones to employ for mostly institutionally demanded elements of our courses (i.e., the need for grades). I’m trying to show our assumptions about our students when we create assessment ecologies that grade student writing and think this is good for them in some way. Do grades motivate students? No, but they do force students, coerce them into doing things, which we usually read as motivation. Coercion is not motivation. Real motivation is doing something even when—and usually when—there is no extrinsic reward, when no one is watching.

And those who say they are motivated by grades I’m guessing have been fooled by the system, integrated into it, thoroughly disciplined, made docile by grades, made to think that their yearning for grades is motivation, because they know no other kind. The yearning for grades, the yearning to be graded, has become so habitual to many students, especially those who have been rewarded by them, that it feels natural. It feels fair and right, but as Freire and Villanueva remind us, part of literacy education is questioning how natural the natural is. Why do I yearn to be graded? How is it fair and helpful to me exactly? What assurances do I actually have that the grades I’m getting will help me in the future?

Grading contracts can offer this questioning because they raise these questions precisely. When grades are absent, many students must ask: What am I motivated by? What do I really care about in the learning, in this work in front of me? At a deeper ethical level, forcing someone to do something because we just know it's better for them by dangling grades in front of them, places ends over means, beyond being a habit of white discourse. I think, both ends and means need to match. Our methods for assessment should match our good ends or goals. It seems deeply disrespectful of others—no matter who they are—for me as a teacher to coerce students into doing things they are unwilling to do, but I can make them because I know what they want. While one can never quite escape a certain level of coercion in classrooms, one can reduce it dramatically. And call it what it is.

Invoking his “liberal role” in “the liberal arts” to justify his pedagogy, Mandel cites a student author from *Change* magazine, who “lambasted her left-wing professors for shooting off their mouths about liberal, human values, and teaching in an atmosphere of stuffy, conservative self-deception.” Mandel continues by quoting the student: “Don’t speak of the liberation of the subjugated and then lower my grade because I hand in a paper late” (630). This is at the heart of the problematic I kept seeing in grades in writing classrooms that attempted to be critical ones or self-conscious and reflective ones. Our pedagogies and writing philosophies are warm, soft, social, and inviting, but our assessment ecologies are cold, hard, individualistic (even selfish), and discouraging. Conventional assessment ecologies are too often uncritical of where their standards for quality and methods come from, who benefits most from them, who has more or less access to them, who uses them to judge, and what the patterned, racialized consequences are because of those judgments and methods, and who controls them.

This blindness or silence in conventional assessment ecologies is a product of whiteness. It is white blindness, or silence, that writing teachers too often have. It is an inability to recognize that we are controlling the rules of a racialized language game that allow us to win, asking (racialized) others to play, knowing that many cannot follow very easily the rules we’ve made that help us win, and by default make sure they lose, because as long as others keep losing, we win. Keep the barbarians in the outer dikes, or better yet, keep them out of all the dikes, goes the tacit logic of the system.

Mandel’s article was the key to convincing me that how much labor students were willing and able to do in my classes was a fairer, more socially just, more antiracist, way to calculate grades. I also noticed at the time that in my classes at Fresno State, an historically Hispanic Serving Institution, most of my students came from working-class Latinx families, many of whom labored in fields in the San Joaquin Valley’s farms, or they themselves worked in the labor

economies in and around Fresno. Bodily labor was important, understood, and I think, respected as honorable work. I became more and more aware of my own body at work and in work, which affected several dimensions of my own writing. Thus it wasn't just the quantity of writing that I wanted to account for in my grading system, but quantity of labor. But this isn't quite true either. What I finally settled on is labor power, a Marxian concept, which I'll discuss in Chapter 3. For now, quantity of writing and time might roughly be equated to quantity of labor, even though this isn't completely true. At this stage, one in which I moved to a labor-based grading contract, I was thinking almost solely in terms of quantity.

Now, one might think that a grading contract system that rewards the amount of work or writing done in a writing class instead of any judgments of quality on the teacher's part would not be very effective at helping the institutionally vulnerable students that I have spend most of my career teaching. I'm speaking of students of color, multilingual students, and working-class students. I'm speaking of students who tend to come to writing classrooms with less contact with the dominant white racial *habitus* that informs the standards in all courses that ask for writing, and that outside contexts and audiences value as most communicative. But in fact, these are the very populations that benefit most from grading contracts. I conducted program-wide assessments of mostly labor-based grading contracts at Fresno State, which showed that most students of color actually do better in many dimensions of quality in final portfolios, and are happier with their writing experiences, feeling that the absence of grades helped them in their learning (Inoue, "Grading Contracts"). Blackstock and Exton also found similar conclusions about hybrid grading contracts in community college and tribal college writing courses, where most of their "nontraditional students also largely fit the profile of 'basic writers' elaborated by David Bartholomae's book *Writing on the Margins*" (279). In their surveying of students, they found that hybrid contracts helped their students because "[b]asic writers perceive that they do not have control over their writing skills, but they can usually control their ability to attend class, engage in academic activities, and complete assignments" (284).

Blackstock and Exton use a similar contract as Danielewicz and Elbow's and say that its shift from focusing on quality of products to the writing process was instrumental in helping their students find the contract helpful in their courses (284). Like all hybrid grading contracts, Blackstock and Exton also found some difficulties with the hybrid aspect of the contract. When discussing motivation, which often goes hand-in-hand with discussions about grades (e.g., if there are no grades, how will students be motivated to revise or turn in a polished draft?), they admit that students who were motivated by such things as grades and percentages will not be satisfied with a hybrid grading contract system. Even still,

they point out that it is important to identify in some way the conditions for an A, the only final course grades in which quality of writing in a final portfolio is considered (288). Still, they see problems and contradictions with so-called clearly articulated criteria for A-quality work in contracts and rubrics, saying that the “language is admittedly subjective and tends to raise more questions than it answers.” So they “are moving toward adopting simple behavior criteria: students may earn an A by submitting additional drafts of the assigned essays” (289). In effect, Blackstock and Exton appear to move toward a purely labor-based grading contract.

And so, this is where my story about how I came to labor-based grading contracts moves to what I use today, a purely labor-based system that only uses quantity of labor, no matter its products, as a way to calculate final course grades. I’ll describe the details of this contract in the Chapter 4, but now I end with Max Marshall, the professor of Microbiology with whom I began this chapter. While he isn’t speaking about grading contracts but of getting rid of grades completely in college courses, Marshall articulates one problem of grades and learning in 1960 (reprinted from an article in 1958) that I think labor-based grading contracts easily solves:

Suppose that, instead of a grade, words are used, perhaps a comment like this: “Slow on the uptake, earnest, reasonably industrious.” This could be equivalent to a “C,” but that student is entitled to his place in the world. He goes to school not to get a teacher’s judgment but because the teacher has more experience than he has. In some situations his talents and efforts may have to be discussed, but this is a sort of indecent gossip and should be restricted whenever possible. To describe his work is bad enough. To use a mere symbol, a “C,” is an unhelpful imposition. The student asks what to do with his sick horse and is told that his horse is sick. A healthy attitude calls for concern with the student’s talents and efforts and what to do with them. A teacher can be friendly with and can respect any honest student. A barrier is erected between the two when the teacher merely bestows the highhanded “C” mediocrity. (Marshall 26)

Words instead of grades. Yes. Learning instead of measuring. Yes. Students finding motivation in and through their own languaging instead of grades. Yes. David Bleich called these things, “descriptive evaluation” (29), and even cites Max Marshall’s 1968 book, *Teaching Without Grades*. Brian Huot called it “instructive evaluation” (69). I think, particularly given the values and understand-

ings about literacies of our field and their connections to writers' subjectivities, we should demand these kinds of assessment ecologies, or least work consciously in ways against the prevailing ecologies of grades.