

Chapter 7. Redirecting Biases in Grading Ecologies

In this chapter, I continue considering labor standards in LBG by focusing on the measures of labor that those standards imply or represent when used to grade or determine completion of assignments. One central concern is understanding the accumulation of biases through the circulation of any labor measure used in a grading ecology. This is something all teachers might consider when designing the measures that their grading ecologies use. I suggest one design process that focuses attention on the biases of the measures used in the grading ecology and discuss the paradoxical nature of such measures in use. I then explore a way to investigate how equitable the outcomes of any given measure may be in a course using my own LBG ecology as an example. Finally, I revisit a statement of mine about labor that has bearing on the biases of labor measures in a LBG ecology, “one hour of labor is worth one hour of labor,” that Carillo is rightly concerned about. This provides me an opportunity to consider the paradoxical conditions we all work and learn in today.

The Accumulation of Biases in Measures for Grading

Regardless of whether a teacher uses a labor-based grading system or not, all grading ecologies operate with their own expectations of labor. In LBG ecologies, labor standards are usually more explicit, numerically defined, and used to determine completion of assignments. This means, as discussed in the last chapter, quantitative measures of labor circulate as quantitative targets for students to complete assignments. I argue that such numerical measures of labor can be easy to understand, negotiated with students, easily incorporated into students’ planning processes in the semester, and less mysterious to students in terms of how they make their final course grades. When these aspects of grading are attended to, students have a better chance at paying closer attention to the noncognitives that predict their success. But quantitative labor measures can also accumulate biases that disadvantage students with disabilities and neurodivergencies, or those with even access to time in the semester.

This problem can stem from labor instructions that offer numerical measures of labor as guides, such as time on tasks and words to be read or written. In the past, I used the number of words produced by a student to determine if an assignment was completed, or if I needed to inquire with a student about what they turned in. This means that the time on task guides in labor instructions can be—should be—a lot more fluid than the number of words to produce in any responses, postings, or other writings. In other words, labor measures that describe time on tasks are *only* guides or estimated times needed to do the work. They are

there for planning purposes and guidance only. The actual time taken will vary for each student, meanwhile words written and turned in are used to determine completion of an assignment by the teacher. The assignments where this may not be the case might be ones where particular responses by the student are the evidence of labor completed, such as addressing three questions in a prompt regardless of the number of words needed to do so. When this is the case, and there's no word count expected, formatting the prompt with a bulleted list of the questions students are to address can help students understand the labor required.

An important part of any labor measure accumulating biases in an ecology is how a teacher responds to any given assignment using that measure. For example, my responses are either (1) full credit or (2) a query and discussion with the student about how they came to turn in what they did. I might ask: "How did you do this labor," or "what happened?" If a student says they engaged in the labor in the spirit it was assigned but the time I estimated did not allow them to produce the number of words I estimated, then they get full credit, and we figure out strategies for them in the future, or I revise my labor instructions and labor measures. If they tell me they didn't spend at least the designated time and that's why they produced less than expected, then it's marked as late or incomplete until they complete the work as expected. I still find this practice mostly serviceable, but it has some concerns, namely what if the student didn't have enough time to do the work in their weekly time frames? What if I was really asking too much too quickly in the term or semester?

Carillo captures this concern about the use of numerical measures of labor, saying that such measures assume "a normative student and a normative sense of time" (16). If how grading measures circulate shapes their biases, then there must be a way to alter these biases or circulate them in ways that make them more equitable. Most of us likely agree that more students today have less time to dedicate to school. Many students' socioeconomic conditions do not afford them as much time as some of their more privileged peers who, perhaps, do not need to work, juggle other obligations, and go to school at the same time. Furthermore, such a single quantified labor standard can be ableist and neurotypical if not enough flexibility in the system exists, if it assumes an ideal way to do that labor, ideal conditions to labor in, and static due dates.

On the other hand, no student who wishes to learn and complete a college education can escape doing work to achieve their development and learning. We, teachers and students, must balance this classroom need for learning with the institutional needs to produce grades and the various limitations and affordances that each student will embody. Flexibility in how labor measures are conceived and used in a grading ecology does not mean that labor standards are so relative that they cannot ever be defined, or that they move and slide around so much from one student to another that the course becomes very different in nature for each student. In such a course, there ends up being no actual expectations of labor time. Anything is acceptable or appears to be.

I also don't think that we can have a course that expects one student to do 50 hours of work while another student must do 100 hours in order to get the same course credit. Keep in mind we are not talking about learning here, just course credit. Learning is always emergent and varied, just as learners and their learning experiences are. Grading, on the other hand, is about meeting institutional demands for credit, and is not an accurate way to measure or offer feedback on learning. Ultimately however, our expectations of labor boil down to hours of laboring. Those labor hours equate to time on individual assignments, time on tasks, even as labor time is not spent exactly at the same time frames, in the same ways, in the same amounts, or at the same speeds for all students in a course.

I doubt any writing teacher would suggest letting students not do work and still get credit for a course. So, how much time in the labors of the course can we expect from our students and how do we determine this? Is it even determinable? The U.S. Department of Education has provided its guidelines; that's the 45-hour per credit rule already mentioned. This seems a reasonable place to begin. It also seems responsible and ethical to be explicit about how many estimated hours a teacher expects from their students; how they've come up with those numbers representing labor time; and how much time they estimate each assignment will take.

Such labor estimates circulate labor measures in ways that provide students affordances. They help students consider their choices and plan the work ahead, even if we all agree such learning and labors will vary across any group of students. Planning is important, especially for students with lots of demands on their time or who may need to think carefully about how much time they need to do work or have in their calendars to do the work of a course. Numerical measures of labor, particularly as guides in labor instructions, help students do this planning directly. Beyond this, it seems unethical, no matter what kind of grading ecology one uses, not to provide estimates and expectations of labor to the very students that have the least room in their schedules for the work of the course.

Our goal then shouldn't be to find inherently neutral measures to use in our grading practices. Those do not exist. The neutrality of any measure used in a grading ecology does not determine its equity in grading. If being neutral and unbiased were the standards for determining what measures to use in equitable grading practices, then we'd have no measures to use. All measures accumulate biases in grading ecologies. So having a quantitative measure that accumulates a bias, such as a falsely perceived neutrality, does not disqualify quantitative measures of labor as ones that might be used to create equitable grading practices. By the same logic, a quality measure is not automatically disqualified from being an equitable measure to use either. Any measure used to grade (or determine completion of assignments) must circulate in a way that provides for equitable outcomes and conditions for all students. A measure that accumulates a bias that assumes the same access to time by all students, on the other hand, is a problem. But it isn't the measure that's the problem, it's the accumulation of biases in or around the measure that's the problem. If the ecology circulates its measures in a

flexible manner, then those measures can acquire different biases, or be neutralized to a great degree.

The central question for a teacher, then, might be: What are the biases accumulated by the grading measures and processes I use and how do my measures circulate and acquire those biases in my ecology? Biases accumulate in a number of ways in grading ecologies. The key in those accumulation processes, however, is in the people, that is, students and teachers. Biases are made by people through what they say and understand about the measures, how those people and the ecology define those measures, and how those measures are used in the system to make decisions and create conditions. I'll come back to these three dimensions of measures in the next section since they help me generate a design heuristic.

We all actually work from these premises already. I doubt anyone believes that even the most experienced and socially just teacher's evaluation and grade of the quality of a student's writing, however you wish to define that construct, is an objective or neutral assessment of that work. Furthermore, no one assumes that the teacher's evaluation is automatically free from ableist and neurotypical standards and conditions. We may trust in that teacher because of their history of socially just judgement, but then we are trusting in what the teacher understands about the measures they use, how that teacher defines those measures for their students, and how they use the measures they do to make the grades they do. We trust the teacher likely because we have some information about their past judgements. We are trusting in what we think we know about how the teacher circulates such measures in their grading ecology. We are not, however, trusting in the measures themselves, nor should we.

Grading Design with Measures in Mind

Working with this understanding about the centrality of the people involved in the circulation of biases, which accumulate in grading measures, our ecologies can define and use their measures in ways that resist, even counter, harmful biases. We can redirect harmful biases that may circulate around our measures if we have such a goal in mind when we design our measures. If I had to think up a simple heuristic to help teachers circulate labor measures in equitable ways, it might start with the following three sets of questions, which explore the three aspects of grading measures that accumulate biases in a grading ecology:

1. **What do I (the teacher) understand about the possible labor measures I can use?** What assumptions and ideas do I have about the measures? Why do I have those ideas about them? What reasonable assumptions and ideas about these measures do my students have and why? What alternative assumptions or ideas might I use in order to make my labor measures as flexible and equitable as possible? How do my students and I make clear

our biases about the measures we use to determine completion of labors or assignments?

2. **How can my students and I collaboratively define the labor measures** we agree to use to determine completion of assignments and final course grades? Will we use numbers, or something else? What can we agree equates fairly to labor and time on task? What kinds of judgements can I (the teacher) make when using these measures? Do my students agree that those decisions can be made from these measures in fair and equitable ways? What processes of negotiation can we use that allow everyone to be heard and have a say in defining our measures for grading? Will we have a future moment in the semester/term to reconsider our labor measures and possibly revise them? Where will our definitions be kept or archived for easy access, reference, or revision?
3. **How will my students and I use the labor measures we decide upon?** How will students reflect upon them periodically? What kinds of measures can students keep on their labor that will allow them to reflect meaningfully on their labor? How will I use the measures of labor to determine completeness of assignments in transparent ways? Will students help in this process in some way? How will I use these same measures to understand the effectiveness, success, or meaningfulness of our grading ecology as a whole? What benchmarks, goals, or outcomes do I think make the ecology successful and/or equitable?

With a little translating, the above grading measures heuristic could be used to inquire about any kind of measures used in a grading ecology, not just labor measures. But there are two important features I'd like to point out. First, the three sets of questions above really are a process for designing an important aspect of the grading ecology: Determining the measures to use in classroom assessment. It focuses attention on the nature of the measures, which then determines what kinds of judgements from those measures are possible. It begins by asking about the possible biases in the measures that will be used to determine grades or assignment completion. The process moves to considering how to define measures with students in democratic processes of negotiation. The last step considers how the measures will be used with students to determine grades and other things, like learning. So the design process is: (1) consider the biases in the measures used, (2) determine definitions of measures, then (3) decide how to use those measures.

The second observation is about the movement of those decisions that is perhaps counterintuitive to many teachers. The movement of this design process determines some important outcomes for the grading ecology that affect measures and the biases that accumulate in them. Because the teacher begins with considering the biases and measures to use in their assessment, those things become more easily malleable. Their construction and definitions are the focus of most of this process. We start with them and end with them.

Now, I don't think this is the typical way most teachers determine such aspects of their grading ecologies. And there is a good reason for it: Pedagogy. My guess is that most teachers, like I did years ago, begin by asking themselves something like, "what are students supposed to learn in this activity? What decisions or judgements do I need to make on this assignment or that one to assess that learning?" Then they ask, "how do I make that decision, or what evidence do I need to produce a grade or determine acceptable completeness?"

Notice the kind of decision that drives this more typical design process: The need to justify an evaluation or grade of an assignment that comes out of pedagogical goals. This can cause problems with the measures used despite the fact that the assessment design process begins with the teacher first thinking about the learning purpose of their activity or assignment, which dictates the kind of assessment decision needed. This places the measures to be used as a secondary concern in the design process. This can mean that their assessment choices are not always in agreement with their pedagogical goals, even though they appear to be at first glance. Why? Because the process doesn't focus on the circulation of biases in the measures used. It focuses on the kind of decision that needs to be made, which is usually aligned with the pedagogical goals, not the biases that circulate with the measures. Remember, assessment is not learning. For these two aspects of the course to agree, the teacher must design agreement in them together.

So a typical assessment design process likely starts with the kind of assessment decision or judgement that needs to be made, which then dictates what evidence or measures are to be used to make that decision. To use a common analogy, the horse here is the assessment decision to be made; the cart is the measure that provides the evidence that will be used to support that decision. In such a design process, the teacher may easily assume the biases in those measures because the focus is on making the right kind of decision, not making the right kind of measure. The teacher can too easily accept whatever the biases of their chosen measures have been for them, or what they imagine those biases are. The important thing is that this more typical design process doesn't afford much room to investigate the biases of measures used because that is not the point of the process. It's much harder to craft biases, manipulate measures, or revise them because the energy is directed at the kind of decision that needs to be made, which again is often a grade or quality evaluation that is dictated by the impulses of our pedagogical goals.

I even question whether most student-driven and collaborative rubric design processes can mitigate this problem. That is, even when students make the rubric, we are not letting them determine the nature of our measures used—either the actual ones in use when reading a given essay or those represented in the rubric—we are letting them determine the constructs that they must demonstrate. Constructs are not measures, but they do often reference them.

Let's say that a teacher has an essay they wish to assess in some way. Since it is meant to serve as a demonstration of learning, and they must provide a grade to

all students according to how well each student has learned, the teacher decides to grade the essay. A conscientious teacher likely will then ask: On what basis shall I determine the grade and how will I explain those decisions to students? Since one of the main learning goals in the essay is to practice using sources and incorporating them into their own discussions, their evaluations will focus on this aspect of the essay, thus it drives the initial assessment decision to evaluate the essay. This also means that the teacher must use some evidence of these practices in their evaluation practice. That evidence should then be represented in the measures they use.

And so, the teacher creates a rubric that has dimensions and descriptions of each grade category possible. In that rubric, they provide examples of the various levels or degrees of effective incorporation of sources in drafts, describing each and perhaps even offering an example or two. In this way, the rubric does two things. Most centrally, it describes each graded essay as a construct (i.e. A-essay, B-essay, etc.). It also references or gestures to the measures the teacher will use to determine grades, likely in those descriptions and examples. In terms of the assessment design process, the rubric is compelled or necessary because of a previous design choice to grade or evaluate the quality of the essay's use of sources—that is, a choice determined by pedagogy with less regard to any socially just assessment philosophy.

The descriptions and examples in the rubric point to the quality-based measures, usually textual features or qualities, the teacher will use to determine each grade. But those measures are usually not the main focus of the rubric. Defining the constructs are. Rubrics describe categories of essays. They cannot describe the infinite ways that students' drafts might be judged to meet such abstract categories. This means that the measures referenced in the rubric are used as static indicators of value or quality if they exist in the rubric. Their natures are assumed to be apparent, at least by the teacher. Why? Because the question asked, the center of the design process, is not about the nature and biases accumulating in the measures referenced. It's about the nature of the assessment decisions that need to be made, that is, what's an A-essay or a B-essay?

In this conscientious design process, one focused on how to assess an essay, the measures are a secondary concern, if they are a concern at all. They are taken for granted as the markers in essays, assumed to be static and understood by everyone. But their meanings and biases actually float, meaning one thing to this student, another to the teacher, and yet another thing to another student. The teacher and students never get an opportunity to investigate the nature of those measures because they are too busy determining the nature of the constructs. Even if the teacher collaboratively derives their rubric with students, the question at hand is still: What does an A-essay look or sound like? What about a B-essay? What features and characteristics will the teacher read for in order to identify each? But this process does not ask: What biases in those features and characteristics do we think will accumulate in our course when we use them to do the work that we have in front of us?

Now, in my suggested reverse design process, teachers might begin with the measures that are available and collectable. It's an inductive-like process, asking something like: Given our goals and work in front of us, what materials and measures do we have available and what do we each know about them? This allows, first, some time to consider the biases and natures of each kind of measure available to use in order to determine some other judgement, which is left open initially. The process then moves to redefining the measures with students. What do we want these measures to signify and do for us? What kinds of decisions seem reasonable to make when we use these measures as our evidence? The focus, now, is on (re)defining the measures, not the judgement to be made from them. This is a conscious move to control the nature of those measures, whatever they are, which ultimately controls the kinds of judgements and decisions made from those measures. Note that after the measures are decided and defined, then we decide on what kind of judgement we want to circulate in the ecology.

As I hope the heuristic above illustrates, a teacher's concerns about the kinds of measures to use in grading should not be just about "which ones should I use or not use." Teachers and students might consider more deeply how whatever measures are used to grade student performances circulate through what we think and say about them, how we define them, and how we use them and explain those uses to each other. For instance, my uses of labor measures are not only about determining completeness of assignments. They are also a way to invite students into negotiations and control over the grading ecology. They are a way for students to reflect upon their labor and the measures of labor in order to understand themselves better as well as control the biases that accumulate in our ecology. Finally, they are a way I can reflect upon and understand better the grading ecology in order to make it as equitable, sustainable, and meaningful to students as possible.

The Both/And of Numerical Measures

What might also be discerned through the above heuristic is that numerical measures of labor are not either good or bad, objectively or subjectively made. They are *both and* perhaps something else. Yes, numbers typically have biases that travel with them that we must contend with, but that doesn't mean they cannot be meaningful and useful in a grading ecology that accounts for these biases and pays attention to them in order to manipulate those biases for more equitable ends. We can change a measure's biases or work with or against them in our grading ecologies. We do this all the time in our daily lives. While most of us agree that numerical ratings can have a false air of objectivity and therefore their use is highly questionable when making decisions on whatever is being rated, many of us use product ratings to make purchases online. We've found ways to make such bias-filled and flawed measures useful to us in our decisions. And yet, they are numerical and can easily be understood as falsely objective measures of the products we are considering.

We typically understand these problems and use such measures next to others with that understanding. We might compare such numerical measures to other measures. What do other customers say in descriptive reviews? What do other websites' product ratings say about the same product? We might even consider the number of ratings and their distribution across the linear scale used. How many ratings actually make up that 5-star rating? We likely read a 5-star rating made from 100 ratings very differently than the same rating composed of 10,000 ratings. We might also consider how many raters rated the product five stars, and how many one star.

This kind of deeper reflection and treatment of such measures is in effect what can be asked of students in LBG ecologies that help them track their labor, keep data on their own laboring, and make sense out of those numbers and words. In turn, these reflections can help students read labor instructions better over time. It's one of the things that I argue is important about LBG in Chapter 3 of the LBG book when I theorize "three-dimensional labor" (106-120/104-118). Such data collection and reflections by students help circulate our labor measures in ways that students can control more. Through structured weekly reflections, students can be asked to consider the ways those measures are highly subjective and contingent. They can be urged to consider how those numerical labor measures are both/and: How they appear to be objective and neutral measurements of labor; how labor guides in instructions compare to their own experiences of labor; and how their own measures are subjectively gathered and what subjective meanings might be made from them. Ultimately, such reflective work by students helps students understand and manipulate the biases that accumulate in the measures circulating in our ecology.

Circulating Measures for Grading

If a grade is connected to rubrics and scoring guides, as in conventional quality-based grading ecologies, evaluations of student performances can be justified analytically. This condition can make students believe that their grades are determined in objective ways, that their grades are objective. In fact, many students may want such grading mechanisms in their courses, since it seems to make clear how their grades are produced. It also seems to clearly show how they are doing in the course. When a grade or points are present, students want to know: How did you calculate that? How was my score determined? Scoring guides and rubrics often directly address these very questions. They answer these questions in ways that address student resistance or concerns, but they also create conditions that present only this kind of question. They frame the evaluative situation, the assessment ecology, as one about a score and how it is calculated, not what did I learn, how did I learn it, and what do I need to learn.

When students are primed with this kind of standards-based condition, then it is easy to fixate on the math that created the grade. And if a teacher doesn't use

a rubric or scoring guide but gives a grade holistically, the problem gets worse for students since they have fewer cues to understand how the grade was calculated. In either case, it's easy to only ask, "how was my grade determined? Where did I lose points?" These are the only questions really available to students since they usually cannot argue successfully to reinterpret the rubric and get a new grade. These alternative questions reveal the biases that accumulate because of the way the measures used to grade circulate in the ecology.

The available questions to students in a grading ecology can also pull attention away from other questions that could reveal the ambiguous, relative, and unreliable nature of the actual judgements used to make any grade. That is, if other questions could be asked and really entertained, then those measures' biases shift. Students, then, could safely ask: Why was my grade determined by these kinds of judgements and measures? Why not other kinds of judgements and measures? Why must that be evidence for your judgement? Why must we use only your (the teacher's) habits of language and judgement to make judgements about my essay? How might I help make judgements on my assignment or labor that count toward my grade?

These questions could reveal that the grade a student receives is just as much or more about the teacher judging, and where they got their languaging, than the student's languaging performance. Bob Broad's work is particularly good at revealing this dynamic in writing programs among teachers, but others have discussed various ways that teachers' judgements say just as much about *teachers judging* as *students writing*. Our judgements of language are shaped in a variety of ways by racial biases (Ball; Matarese and Anson), implicit biases (Banji and Greenwald), false assumptions about objectivity and neutrality that amount to racist judgements and assessment (Randall; Russell), disciplinary and ideological considerations (Faigley; Anson), and the activity of judging itself that makes it inconsistent and idiosyncratic by nature (Belanoff; Deiderich). And I'm not even referencing the ways teachers construct "error" in vastly different ways that tend to harm language minoritized students (Williams; Horner; Anson).

What further compounds the above problems with quality-based measures and judgements circulating in grading ecologies are tacit or unexamined expectations of labor. Not using quantitative measures of labor to determine grades in a grading ecology does not mean that labor expectations are not present. They always are. We always expect students to do labor. And how we understand that labor and expect it from students is an equity concern.

Labor, the actual work done by students, can easily be forgotten or unaccounted for in assignment expectations and grading. I mean, if you aren't going to use measures of labor to grade, then you likely will focus most of your attention on other measures, like quality measures. In such cases, only the teacher controls quality measures and what they mean for a grade. When quality measures circulate in grading processes, they are a function of the teacher's judgements of language. The fact that the students in the course are laboring in uneven ways has

little to no bearing on those processes that produce grades. And yet, as Carillo reminds us, such unevenness in labor across students in a course are key sources of inequality in grading ecologies, particularly when we can expect to have students with disabilities, illnesses, or neurodivergencies. And so, we all should be asking: How are the measures we use to grade in our grading ecologies accumulating ableist and neurotypical biases around labor expectations?

In practice, what makes any measure, including numerical measures of labor, difficult to use in equitable ways for grading is that beyond using the measures to determine grades in a course, we also need to communicate those measures in unambiguous ways to students in assignments. And so, there are two problems when circulating labor measures. There is a problem with the use of numerical labor measures in labor instructions as estimates or guides for students, and there is a problem in their use by a teacher to determine assignment completion, which ultimately determines final course grades in LBG. We might then ask: (1) Should the ecology provide quantitative estimates or guides of labor to students, say, in labor instructions; and (2) should the ecology use those guides or estimates to determine completion of assignments? I think the first is less of a problem than the second can be, but the questions are bound together.

Defining tasks and assignments by numerical measures of labor (e.g. time on task and number of words read or written) can help students understand what they are being asked to do and how much time they need to do it. The estimates don't have to be completely accurate or spot on for every student in order for them to be useful to any student. For example, each semester I open discussions about reading labor instructions by talking about how it is impossible for me to determine universally accurate labor estimates in my instructions. In my experience, students usually understand this natural unevenness. They understand that my guides in instructions cannot be perfect measures of how much labor to do because we are all different. We ask: How shall each of us translate the labor guides? How should we read those labor guides so that we can do the laboring in meaningful ways?

One way I have explained to students how to translate my numerical labor estimates in labor instructions is as middle of the road guides. This explanation helps me validate or check the fairness and equity of my estimates after a course is over. I tell students that my labor guides of time on tasks in instructions are meant to identify where I initially think the mean or median student in the course is located. Numerical labor measures in instructions, then, define what I've estimated the student in the middle of the classroom likely needs to accomplish each assignment.⁷ I craft such estimates by trial and error from past courses and students' efforts, as well as paying careful attention to their labor logs, tracking documents,

7. I do realize that the mean (average) student labor in a class is not the same as the median student labor, but for a discussion of labor with students, both work. I tend not to say the "average" students, as that term can be heard as a norm or even be normative.

and journals. If I've accurately done this, then it should be safe to assume that most of us will need either more or less time than what I've estimated, but not a lot more or a lot less.

Labor guides for time on task, then, do not have to act, nor represent, a "norm" for labor time spent in a course, at least not a heavy-handed norm everyone must follow precisely. Everyone varies in their laboring. These measures can instead be an estimated mean, which has been derived from past courses like the present one and will continue to evolve. Therefore, a teacher would be wise to make clear that most students take more or less time than the guides can represent. The biggest difference, then, between labor guides that are normative (in the negative and oppressive sense) and ones that are an estimated mean, is in how they are defined, discussed, and used by students and the teacher, which I discussed in the previous section, "Grading Design with Measures in Mind."

As Appendix B illustrates, a teacher can do this work in another document that students read in order to begin discussions of labor in the first week of a semester. This is one way I've found to address the three kinds of questions in the heuristic from the previous section. The appendix offers my own version of a past course's "Defining Labor Document," which introduces students to the course, our labor-based grading ecology, and how they'll plan and track their labor through the semester. The document also explains how I've determined our initial labor estimates and how they might begin thinking about translating them in labor instructions. This is their very first assignment on the first day, which is a reading and reflection activity.

Of course, the proof of success of equitable measures of any kind is in the outcomes and details of a course. What happens in a LBG ecology when labor is defined, discussed, and used in the ways I'm suggesting? We might consider a courses' distribution of total labor recorded next to labor time estimates. In particular, consider the dispersion of those values (i.e. the total labor logged by students) on a graph. That dispersion can be represented by the standard deviation (SD) of those values, or how dispersed the values are from the mean value in the data set. In this way, SD tells us how far apart students were in their total amount of labor done in the course, which is a function of how they translated numerical labor guides and the tasks given in labor instructions. If the teacher has been too erratic or asked too much and not estimated enough time in those instructions, then the graph may have a higher (worse) SD. The graph will show more dispersion. The graph may also show too many students doing a lot more labor than was estimated. This could show as less dispersion with too many students grouped in a part of the graph that shows more labor time logged. Finally, if too many students do worse grade-wise than the mean student, then there may also be a problem with the grading ecology, regardless of how the SD turns out.

In a recent typical course, I estimated a total of 118.67 hours, or 7,120 minutes, of labor in the semester in all of my labor instructions. That's about 12 percent less

time estimated than the 135-hour rule. If we assume a normal distribution of labor time by all students in the course and its “empirical rule” of 68 - 95 - 99.7, then we can check how close my estimates were for that group of students.⁸ If we accept these rules for the distribution of labor values, then about 68% of the students should be within one SD from that mean on either side (i.e. logging more or less labor than the mean), 95 percent of all students should be within two SD, and 99.7 percent should be within three SD from the mean.

In this recent first year writing course, I had 15 students who completed their labor logs. The mean (or average) total labor was 6,365 minutes logged. This is just about 12.5 hours (755 minutes) of labor less than I estimated (about 11% less labor). The SD, calculated by Google Sheets STDEV function, turns out to be 1,380.80, which makes the boundaries for one SD at 7,746 (high end) and 4,984.40 (low end). If my students fit within the empirical rule then around 68% of them should be within these boundaries.⁹

When all 15 students are plotted on a graph, as in Figure 1, eleven students are within one SD from the mean. That’s 73 percent of the students. Three students are high, logging more labor, and one is lower, logging less labor, but all are well within two SD from the mean. No one sits three SD from the mean. This means that no one has done a lot more labor or a lot less than what I estimated. Keep in mind that in typical grading curves, three SD from the mean tend to be grades of D/F and A/B, whereas two SD are roughly D/C and B/C.

The four students who are between one and two SD from the mean were all White students, while the six BIPOC students in the course were all within one SD from the mean. Additionally, the mean of the class was 6,365.2 minutes of total labor. My estimate was 7,120 minutes of total labor. This means I estimated about 11 percent more labor time than what the mean of this group of students needed to complete the work. Keep in mind that I’m not asking students to do the exact amount of labor I’ve estimated in labor instructions. The numbers are general guides that need translating. So, we have to expect a range of laboring, some kind of dispersion, and how much dispersion might be one way to consider the equity and fairness of the circulation of any labor measures. In the end, I consider this a pretty tight dispersion. Given that the actual mean is lower than my estimated mean, and my estimated mean was lower than the federal guidelines for courses’ labor hours expected, it can be argued that how my labor measures circulate create equitable conditions in my course and this accounts for an inherent amount of difference in laboring.

8. The empirical rule is the typical way most understand the convention boundaries of “normal distributions,” or bell curves. Not all values in a range will do this, and one could argue against the validity of the normal distribution, but it is one way to check the dispersion of a range of values.

9. There is a good definition and discussion of standard deviation by Paresh Khandelwal that I feel those with little statistics background can read.

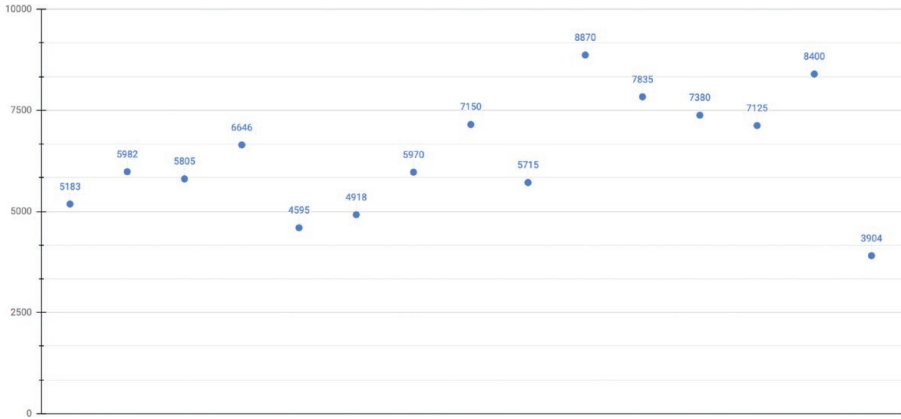


Figure 1. The total labor logged by each of the FYW students in a recent course are all within two standard deviations from the mean.

As I tell my students, I generally overestimate how much labor time I think is needed in all labor instructions by about 15 percent. I also explain that I try to assign a total amount of labor that is 15 percent less than the 135-hour guide (that boils down to about 115 hours). So as long as I'm estimating 15 percent more labor in instructions and 15 percent less total labor in the course, then I think I am in safer, fairer territory, than if I was trying to be perfect. That is, if a teacher estimates too few labor minutes in instructions, then the average student would need more time to do what I've asked. This risks asking too much work of students in the semester. In the above example, it does not appear to be the case. Most students (nine students) generally took less time than estimated to do the work, but six students took more time, with two of those students only marginally more time. Two of these six were BIPOC students. Thus, the tight dispersion of total labor logged, with no BIPOC students doing a lot more labor than estimated, suggests to me that there's no undue effect on any students in this course, despite the fact that I cannot know for sure all of the students who may have disabilities or who experience neurodivergencies, since many do not officially claim it and I cannot ask about it.

This kind of analysis of labor log or tracking data is encouraging to me, as it suggests a fair and universally accessible grading ecology, not an ableist or neurotypical one. Now, one could argue against the bias of the bell curve that is assumed to be okay in my discussion. In fact, this course's dispersion does not appear to be quite bell curvy enough to be a true bell curve. This doesn't invalidate the way I've understood the variance in the ecology, though. For example, there are no students between two and three SD from the mean and most students hover above the mean of the course. Then again, I'm not interested in replicating a perfect bell curve. I'm only interested in variance and how close students are to the mean—that is, SD. Furthermore, these features could be due to having only

15 students in the course. Most statisticians suggest at least 30 in a sample size to meet the Hungarian mathematician George Pólya's "central limit theorem," which states that a large enough sample size will give you a normally distributed set of values (Ganti). Then again, this isn't a sample analysis. It's all the students who finished the course.

I should note that I have argued against what I call "the bell curve bias" in grading systems (Inoue, "Why Does Conventional" n.p.). This is a bias we all have and it comes out of enlightenment traditions that are closely linked to eugenics and racist logics for ranking people in all kinds of ways. When we have such a bias in systems like grading ecologies, then we often judge the effectiveness of our systems on how well the final grades in a course match a "normal distribution" or a bell curve. This is dangerous and can lead to racist outcomes.

My objections to the bell curve bias are mainly centered on using it to validate the final distribution of grades. I do not argue that we cannot find bell curves in nature and even other phenomena. In fact, all we have to do is consider Galton's quincunx, or "the Galton Board," which demonstrates his theory of "regression towards mediocrity" or the mean (Inoue, "Why Does Conventional" n.p.). Through the random falling of beads into a matrix, a natural bell-shaped distribution will occur if enough beads fall.

In my case, I use a bell curve bias as a way to understand the distribution of total labor occurring in my course. I do not use its logic to guide how grades are calculated or distributed in my course. Remember, I'm only considering the labor logged and estimated in my discussion above, not grades that my students ended up getting. I am trying to understand variance first as a key element to fairness and equity of labor estimated and done by students. It is this variance that is at the heart of equity concerns for students with disabilities, time constraints, and neurodivergencies, and it is activated by the accumulated biases in labor measures in a LBG ecology.

But what about the grades in this course? How did these students do? The data above suggest that the use of quantitative labor measures can function equitably in a grading ecology, at least if we accept how students' total labor logged was dispersed. The grade distribution of this group of students provides a fuller sense of the ecology's equity. Everyone in the course received a B or higher grade, most getting A's (12 out of 17). Additionally, all five students who received Bs did less labor than I estimated, except 1 student (a BIPOC woman). She clocked in 7,150 total minutes of labor, while my total estimated labor was 7,120 minutes. She was right on target. And since this course's grading contract was for a B course grade, these numbers seem fair and equitable.

In the course evaluations, in which 11 of the 17 students who took the course responded, several students (4 of 10 responses) raised questions about the amount of work asked, but the course grading structure was still rated highly, averaging 1.4 out of 5 with 1 being the top rating. Most students who responded also acknowledged that the course was a 7.5-week, asynchronous online course. It was

expected to be a course that asked the same amount of work in half the time as a typical 15-week course. Thus, the concerns students' raise about workload was not about it being unwarranted, it was about the length of the course, which most admitted they'd signed up for knowing beforehand.

Ultimately, numerical labor time estimates in LBG do not have to circulate with normative, ableist, or neurotypical biases. They can circulate as simply numerical guides for labor time, even if other numerical measures are used to determine completion of assignments, such as number of words written and due dates for some time-sensitive assignments. One key in the above example is to overestimate time on task while being careful not to ask for too many words. Another key is to estimate a total amount of time in the semester that is less than generally expected, such as the 135-hour rule. Finally, flexibility in due dates of assignments can be built in so that late work does not count against contracts, or it is greatly reduced in its impact on final course grades.

“One Hour of Labor”

In Chapter 6 of the LBG book, I provide three measures that guide my decisions for completed labor, and they identify the three main measures I originally found meaningful:

- Is the labor product(s) posted on time and in the correct place?
- Does the labor product(s) include everything I asked for and meet the minimum word count?
- Is there a labor tweet/Slack(s) posted as instructed (if applicable)? (200/197)

As I see it, the first two measures are key. They can accumulate unfair biases in an ecology that uses the above three measures. Because I did—and still do—have concerns with how measures of labor are used in LBG, I qualify this three-item guide for how to count labor done for contract purposes in the book. Here's how I end the same section:

In the end, much of what I count as complete labor is done by trusting my students and done in as quantifiable way as possible, always trying to give the student the benefit of any doubts I may have, even if I may still ask that student about their labor if it seems to be less productive that I hoped for. (203/199)

While this is not enough to fully address the problem of ableist and neurotypical standards of labor, it should illustrate that I believe we must trust our students, listen to them, and not be overly strict in applying our expectations of labor as measures for completion of assignments. It should also mean that our students can help us design our measures, as my design heuristic in the previous section of this chapter implies. When I find a student who appears not to meet

the labor expectations based on the above measures, I talk with them, try to learn what has happened, and give them the benefit of any doubts. This is one version of flexibility that I hear Kafer and others promoting.

I admit that at the time of my book's original drafting, I was thinking that the more quantifiable I could make labor, the fairer I could judge that labor as complete for grading purposes. I was most concerned about making a grading ecology that was antiracist, that countered the White language supremacist systems that were created by Whiteness language standards and judgements of quality, and not necessarily one that countered ableist and neurotypical standards of labor. In her criticisms of LBG, Carillo rightly points out that I assumed that labor can be a better and more equitable measure of learning in a grading ecology than judgements using a single standard of quality (9-17). I still believe this, even if those measures of labor are not just paradoxical but problematic (in the Freirean sense).¹⁰ Thus my understanding of this is not without its tensions, as I hope you can hear from my discussion.

Perhaps my choice of words in the phrase "one hour of labor is worth one hour of labor" casts too long a shadow over my discussion. Carillo quotes this clause and uses it to argue that I use my numerical labor measures as falsely neutral measures, as measures that I use as if they don't have biases that can harm students with disabilities or neurodivergencies (11). I do appreciate this criticism and accept that my statement is misleading and incomplete. Even when I wrote the LBG book, I did not believe labor measures were neutral. I did and do still believe, however, that they can be safer and more equitable measures than quality measures.

My original full sentence, however, does attempt to capture some of the paradox in labor measures that I think resists a one-sided reading of my sentiment: "One hour of labor is worth one hour of labor, regardless of the kind of labor you are engaged in during the hour and even though not all labor is equal when understood in terms of other domains, such as learning or engagement" (131-132/127-128). This statement along with my conclusion with which I started this section support the idea that we can address, and perhaps avoid much of the harmful biases that often accumulate in numerical labor measures, if we listen to

10. In his discussion of problem-posing education, one that encourages "critical consciousness" or "*conscientizaçao*," Paulo Freire explains that there is a dialectic between people and their situations, that is, "[h]uman beings *are* because they *are in* a situation. And they *will be more* the more they not only critically reflect upon their existence but critically act upon it" (109). This reflection and action dialectic leads the individual to the "objective-problematic," or an understanding of oneself as both of and in a situation or set of conditions (109). Freire's problematic recognizes the consubstantial and intersubjective nature of our world and ourselves, thus when I say "measures of labor are problematic," I mean that they are measures of both the individual and of the conditions in which that individual labors, which mutually constitute each other. And it is in this problematic that students can come to understand labor in meaningful and educative ways.

and trust students, if we build flexible ways to circulate labor measures together. At the same time, I consider all kinds of labor (e.g. reading, writing, researching) the same in regards to the grade breakdown table in the contract, and doing so asks us to engage in all the labors of a course with equal care (132/129), while acknowledging the fact that labor is unevenly done and experienced by students.

I find Lorena Gibson, Grant Otsuki, and Jordan Anderson's summary of the way labor circulates in LBG to be another good way to explain what I'm trying to say here. All three teach cultural anthropology courses at Te Herenga Waka - Victoria University of Wellington (in Aotearoa New Zealand). They frame LBG as a way to address the concerns of indigenous students (mostly Māori and Pasifika) and other students that became even more prominent during the early months of the COVID-19 pandemic and lockdowns that happened in Aotearoa New Zealand. Drawing on Nel Noddings' feminist work and my LBG book, they explain LBG as a "practice of care." They summarize and rearticulate my theorizing of labor in LBG:

In conventional classrooms, grades circulate as a primary unit of exchange [2019: 81]: students produce pieces of writing which act as commodities that they exchange for grades. These grades can then be used to acquire other valuable things like entry into a college, a scholarship, or a degree. Insofar as it is the *writing* that is exchanged for grades, conventional grading systems teach students (and instructors) to *care about the written product*. This reflects a neoliberal assumption that equality among students is based in equalising their opportunities. Then, the value of their contribution should be measured by how they maximise the return on their investment in that opportunity, and their grade should reflect the size and quality of that return. In contrast, the objective of LBG is to make students' *labour* rather than their writing the valued commodity. LBG equalises the value placed on the units of labour they devote to their writing, such that their task (and ours) becomes carving out a structure and space that allows them to do that labour. In other words, LBG shifts the locus of value in the classroom so that students and their teachers come to *care about their labour*, and by extension *care for the labourer*. (41)

I realize there is a paradox in this understanding of labor. While one hour of labor is always one hour of labor numerically, what that hour produces for any given student is not the same across any group of students. One student's hour may also be experienced very differently than other student's hours of labor. One hour of labor may also be more or less accessible to some students than others. A grading ecology simply cannot control these variables, which I think also means we shouldn't use students' experiences of labor as a measure to grade either, even

as we need them to help us reflect upon that labor and interrogate the biases that accumulate. And so, I wonder: Is using “engagement,” arguably a very phenomenological aspect of labor, an ethical measure to use in a grading ecology? I’ll say more about this in Chapter 10.

For now, I believe my best course of action is to lean into this understanding about the unevenness of the experiences of laboring, work with this knowledge by working with how labor ends up getting dispersed. We can mitigate the effects of the uneven ways students labor by having more flexibility in due dates and reconsider whether to count late assignments against a student’s contract. These are good changes to my original system which I describe in Chapter 11, and, as Gibson et al explain, they can be understood as a part of an ethic of caring for the student through caring for their labor.

This still leaves those word counts as the primary measures that decide grades, at least facially. This means that the more normative guides in instructions are the word counts. That marker, how many words are turned in, signals whether I need to talk to the student or just respond to their assignment. Everyone’s labor is equal in this way. But is this equality without equity? Is it equitable to assume that one student’s 200 words submitted is roughly the same amount of labor time as another’s? Not likely. Does each hour that each student spends, regardless of who they are or their life circumstances, count exactly the same in terms of learning? Not likely. But in terms of their grade? I think it has to, as paradoxical as such a practice is. I recognize that this tension is not adequately addressed in that “one hour of labor” sentiment.

I don’t see a way around this unevenness in what words submitted represent in terms of time. A teacher might ask students to spend time only on assignments and let the amount of words fall where they will. But how would that teacher know with any accuracy the amount of time in minutes each student spent on each labor. Teachers might trust their students more, and this is surely a trust issue for me, but it would also affect students’ labor tracking and reflections, making those activities much more about accounting for their labor, not reflections upon it to make meaning out of it. That’s more important to my LBG ecologies at this moment. I don’t want to create the appearance of counting minutes in labor logs or tracking documents and turning those reflective documents into surveillance, or an accounting system in service of their grades. So, I prioritize the numerical measure of words submitted as a substitute for labor in the ecology that offers student control, reflection, and flexibility with due dates and time expectations of labor.

In the past, when students have had trouble with the time estimates not equating to the number of words expected, I suggested that time should be the primary factor they use to guide them and let everything else be what it is. But this is an explicit agreement with an individual student after we’ve talked about how they do their work in the course. I let them know that I will account for this when I get their work, and I do. Another option is to let students decide which labor

measures to use as their primary guides for any given labor goal. They would pick the preferred labor measure to be used to determine completion of each assignment and tell the teacher when they turn in the assignment. They would explain why that measure better reflects the spirit and goals of the assignment. Again, a great amount of trust in students is required here, as well as extra time and agreements about how to explain labor measure decisions by students would be needed. Such things may not be understood well by outside stakeholders of the course.