Response to the EAC report on Teaching Load Guidelines

Priscilla Bremser
June 15, 2017

I. Introduction

Since 2010, when the then-EAC (Educational Affairs Committee) began sharing draft proposals for new teaching load guidelines (TLGs), members of my department have had misgivings about them. We identified both structural weaknesses and potential repercussions that were not being acknowledged. When we've raised our concerns, we've been told to wait and see, and that the EAC would analyze the relevant data after a few years. Now that we have the EAC’s May 2017 report, I offer this response, in hopes of well-reasoned dialogue about the “value of the TLGs as a measurement tool” (page 9 of the report).

I welcome the EAC’s call to “decide on a way forward that allows us both to measure teaching loads accurately in the interest of equity, and to focus on what is truly important in our teaching” (p. 9), while cautioning against unrealistic expectations for using quantitative data to produce equity. We offer our students a rich array of courses from dozens of departments and programs in a system that offers one graduation credit per course.¹ Our colleagues direct creative writing workshops, work with students in laboratories, extend office hours well beyond the three-hour-per-week minimum, guide aspiring artists and musicians, and grade stacks of twenty-page papers. Many of us supervise independent and thesis projects that are unrelated to our own research. We simply cannot expect a single metric, or even a conglomeration of several metrics, to capture all that we do.

This response has been informed by many conversations with colleagues. I am not an applied mathematician, much less a statistician. Bill Peterson (trained in Probability) and Albert Kim (Statistics) have been particularly helpful as I try to articulate my concerns.

II. Concerns about the guidelines themselves

Early in 2016, the EAC asked department chairs for feedback on the TLG’s. As I wrote at that time, the major challenge for our department, in both theory and practice, is that these TLGs were built on faulty data analysis. As a result, the goal that every faculty member be at or above average in at least two of the three categories is numerically impossible to meet.

¹ There has been some discussion of changing to a credit-hour system. In this response I assume that we will retain the current credit system.
First of all, the stated norm of four contact hours per course prep was chosen by using the mean across all courses, rather than the more appropriate median. But most courses - 57%, according to data shared by the Dean of the Faculty - meet for three hours per week (thus our course schedule template)\(^2\); it's a relative few at the high end that pull up the mean. Hence most faculty members are out of compliance on this metric.

Similarly, using the average number of course “preps” ignores the distribution, and puts anyone who teaches multiple sections of the same course (which makes pedagogical sense in some departments) out of compliance. Having outliers on the high side will pull the averages (means) above the medians. It seems reasonable to conclude that, college-wide, we do not have majorities meeting the target in any category. This conclusion is confirmed by the first three figures in the EAC’s information report document for the March 17 meeting this year.

Had the guidelines been based on medians rather than means, however, the expectation that “faculty who are below the average in one of these measures by any significant amount should, at a minimum, be at or above the averages in both of the other two, and, to a corresponding extent, above the average in at least one” would still be impossible to meet.

To see why that is the case, imagine a faculty with just eight professors. By definition, only four of them can be above the median in each measure. The chart below shows one possible state of affairs. A “1” means the professor is above the median in that measure; a “0” means the professor is below.

<table>
<thead>
<tr>
<th>Professor</th>
<th>Contact Hours</th>
<th>Preps</th>
<th>Enrollments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor 1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Professor 2</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Professor 3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Professor 4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Professor 5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Professor 6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Professor 7</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Professor 8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

| TOTAL AT OR ABOVE AVERAGE | 4 | 4 | 4 |

Because, by definition, half of the professors will be below each median, there will be four 0’s, and hence four 1’s, in each column, for a total of twelve 1’s. But the goal

---

\(^2\) Here “three hours per week” means three 50-minute or two 75-minute sessions.
is to have two 1’s in each row, which would require sixteen 1’s (two-thirds of the total of 24 scores). No matter how the twelve 1’s are distributed among the cells, there aren’t enough for each professor to have two of them.

Now imagine a chart with 250 professors. Each of the three columns would have 125 1’s and 125 0’s, for a total of 375 1’s, not nearly the 500 needed to put two in each professor’s row. Again, this problem is made worse by the insistence of the EAC on using the mean rather than the median to set benchmarks, since the means tend to be higher for these three measures, as shown in the EAC’s March report.

The logical fallacy of the “two out of three” construction (the Lake Wobegon clause) should have doomed these guidelines from the start. We can’t be at or above the average in more than the average number of categories. Since the guidelines are still in operation, however, I next identify my concerns about the measures themselves.

Most obviously, the guidelines create incentives for large class sizes, which is pedagogically questionable and certainly inconsistent with how the College markets itself. This is a consequence of setting an “average” target for each metric and then requiring offsets in one category for shortcomings in another. More generally, the guidelines create incentives for a variety of decisions that make no pedagogical sense. A senior mathematics major figured this out instantly when we explained the difference between a “course” and a “prep.” He said, “they want Bremser and Dorman to each have one section of Calc I and one section of Linear Algebra, instead of each teaching two sections of the same course? What’s the point of that?”

Perhaps less obvious is the fact that the guidelines privilege certain kinds of teaching over others. For example, a single instructor teaching a course with a 3-hour lab gets double credit: both added contact hours and an extra .5 prep credit. An instructor whose course has labs run by an ASI (assistant in science instruction) gets credit for all of the students in the course, even though there’s another scientist instructing those students, assessing their work, and holding office hours. We recognize that well-planned and supervised laboratory sessions are essential to science instruction. But the TLGs do not offer parallel recognition to pedagogy-driven work patterns in other disciplines.

Having spent three years on the Committee on Reappointment, three on the EAC, and five as department chair, I am quite confident that the vast majority of Middlebury faculty members are pulling their own weight. It is important to acknowledge the ways in which we stretch ourselves beyond averages, as dictated by the needs in our own programs. It is unnecessary and counterproductive to convert that acknowledgement into penalties for colleagues in programs with different needs.
III. Concerns about the analyses in (and not in) the report

In preparing its report, the EAC states, "we developed a fondness for a hybrid model whereby the current TLGs are kept or adjusted," as opposed to a complete overhaul. However, the report fails to justify that fondness.

In January 2011, the then-EAC offered “three reasons why we are moving away from IUs. First, IUs have caused us to focus too heavily on enrollments in our course planning. Our plan is to move to a system that allows us to focus on pedagogy and learning goals. Second, the EAC recognizes that no system can ever fully reflect our workload. Although no system is perfect, the EAC wants to focus on a system that is fair enough for everyone and allows us to concentrate on pedagogy. In fact, IUs have served as a proxy for the three measures that we want to focus on: student enrollments, courses and contact hours. The new system will consider all of these things. Finally, none of our peers use IUs which makes it difficult to compare our teaching loads.” It seems reasonable to expect an EAC report on TLGs to judge them on the basis of those three purposes.

The current EAC admits in its report that the first objective of allowing faculty to focus on pedagogy and learning goals has not been met. From page 8, “we regret that the TLGs have invited this focus on accounting and its attendant anxieties, and would like to see us all move away from a focus on ‘what is counted how’ and toward a focus on what is pedagogically best for students.”

As for fairness, the report offers a collection of charts and graphs, followed by this statement on page 7: “(t)o summarize our quantitative examination of the data, then, we found it largely reassuring.” I found it quite troubling. The EAC’s “quantitative examination” does not actually address the question of equity. For one thing, the formulation of the guidelines relies on colloquial misconceptions about the meaning of statistical terms, misconceptions still at play in the latest EAC report.

First, the term “average” sounds like “typical” or “what most people are doing.” In that sense, asking colleagues to be at or above the average in two out of three categories sounds reasonable. In fact, it turns out to be arithmetically impossible, as shown above.

To dig a little deeper into why this is the case, note that our distributions tend to have natural lower bounds--say 3 contact hours for a course--with outliers predominantly on the high side. In such situations, more than half of the courses will meet for less than the target of 4 hours. Since we are averaging our teaching experience over 4 years, some of the asymmetry will be mitigated. But it is simply not possible that a majority of colleagues will be at or above the average.

-----------------------------------------------

3 Minutes, Faculty Meeting of January 10, 2011.
In fact, that is exactly what was shown in the first report issued by the EAC. As Bill Peterson pointed out at the April faculty meeting, in each category, the median performance was at or below the target. This means that 50% or more of the faculty are at or below the target. It is therefore impossible that everyone is at or above the target in two of three categories.

To clarify that point, Bill asked the EAC in April (and on several previous occasions) to produce the distribution of the number of colleagues at or above the guidelines in 0, 1, 2, or 3 categories. This is the most natural question to ask, given that the guidelines set targets for individual faculty members, not averages for all at once. It is also easily answered from the data already collected. I was mystified, frankly, that the EAC didn't include it in the final report.

The second misconception lies the term “proportional,” which is colloquially interpreted as some kind of fairness. In fact, there is no logical reason to suppose that a 10% shortfall, in say, preps is fairly compensated by a 10% increase in contact hours. But this is exactly the tradeoff proposed in the Guidelines. Such thinking is further reinforced by the “average of ratios” introduced by the EAC in its latest document, with the dubious claim that this allows an “apples to apples” comparison of teaching loads.

It is ironic that the EAC has suggested that colleagues are allowing themselves to become preoccupied with the guidelines rather than focusing on pedagogy. There is no pedagogical reason to consider gratuitously adding contact hours to an existing course to make up for a perceived shortfall on some other measure.

Nevertheless, the latest EAC report is fixated on the “average of ratios” idea, devoting at least a dozen tables and graphics to it. But the targets were based on averages of what we were doing. To say we are on average meeting those averages is almost tautological; it certainly does not require volumes of color graphics for support. It also contributes nothing to a consideration of equity in teaching loads.

Because the TLGs set expectations for individual faculty members, it would have been more helpful had the EAC presented the distributions of the data for each measure, using the same units as the TLGs. In a recent TED talk, the data scientist Mona Chalabi suggested the question “Can I see myself in the data?” as a way to evaluate a statistical argument. Can individual faculty members see themselves in any of the bar charts that display the newly-invented “Preps Ratio,” “Contact Hours Ratio,” or “Enrollments Ratio”? Have any individual faculty members computed those ratios for themselves?

Further, the “average of ratios” approach introduces even more incentive to teach large classes, pedagogy be damned. For just one example, I heard from a colleague in

---

4 https://www.ted.com/talks/mona_chalabi_3_ways_to_spot_a_bad_statistic
a different department who is thinking about weakening prerequisites in an upper-level course to appeal to a larger audience, thereby cheating the more experienced students out of an appropriate challenge, just to get that enrollment count up. The enrollment count is the only changeable measure; the other two levers are stuck for that instructor.

What’s also missing from the EAC report is an accounting of the consequences of the double credit for labs mentioned above. For example, in the 2016-17 year, there were at least 31 courses with enrollments at or under 24 in which the instructor led a three-hour lab. In addition to logging six (or more) contact hours, each of these instructors got credit for 1.5 course preparations. Thus an extra 15.5 course preparation units, the equivalent of 3.4 FTE’s, were allocated to those departments. In an austerity environment, it would have been useful to get a complete count of the FTE cost of those extra prep credits.

Wisely, the current EAC chose not to mention the 2011 goal of simple comparison to teaching loads at peer institutions. Trying to explain the TLGs to mathematicians and statisticians at conferences is a particular challenge.

IV. Moving Forward

As economists know, sunk costs can prevent humans from making rational decisions about the future. Nonetheless, I suggest to administrators and faculty colleagues that we learn what we can from this TLG experiment and then move on.

One lesson I’ve learned from EAC reports is that faculty and departments appreciate flexibility in planning their teaching over multiple years. That flexibility has also allowed us to see that when given a chance, many faculty members move away from teaching Winter Term.

As suggested above, there are further analyses that could be done with the teaching data the EAC used in its report. I urge the EAC to share the data with colleagues who request it. Crowdsourcing could inform our deliberations.

Recently I came across an interesting problem for middle-school mathematics students. Given a set of scores for several components of an imaginary game, students were to “invent a scoring system where your team wins.” Any of the teams could win with a cleverly chosen system. This crystallized the essential TLG problem: we have been expecting our teaching load guidelines to do the impossible. Naturally, we want a mechanism to ensure fairness. Our teaching work is so complex and varied, however, that there is no neutral way to capture it all in one “average of ratios” or the like. One team or another will likely win. That is, one set of teaching practices will be favored by what “counts,” to the detriment of disciplines in which different practices are more appropriate.
The current guidelines are an attempt to build a mathematical model—a mathematical representation of a real-world situation—to measure our workloads. Applied mathematicians like to say, “All models are wrong; some models are useful.” My children didn’t grow exactly as the growth charts predicted, but close enough to prevent worrying on that front. The current TLG model, on the other hand, is simply not useful if the EAC suggests that the guidelines be “used mostly in the background” (page 9).

As we're all well aware, the model is built on just three components. That senior math major also asked, "do office hours count?" Office hours in Warner were apparently a central part of his experience. We can all list components of our work that are valuable to student learning but don’t "count" in the TLGs. Indeed, some of them are listed in the EAC report.

Every choice of a component, and the weight given to that component, is a value judgment. This is one reason the guidelines provoke such anxiety. Why three-hour labs and not one-hour labs? Why a 1.5 weight and not a 1.25 weight for that course with a three-hour lab? Why not CW courses? It’s also why we will never agree that a particular system of this type is “fair.”

Further, because the targets for the three components were based on averages, we’ve taken the special challenges of large class sizes, extra contact hours, and multiple course preps and turned them into virtues. We should find ways to acknowledge and perhaps compensate for those challenges without pressuring others to artificially construct them for themselves.

We should also have guidelines that we can easily explain to each other and prospective colleagues, and compare to those at peer institutions. I understand the EAC’s resistance to an overhaul, but the proposed alternative—pay no attention to those guidelines behind the curtain!—is untenable.

Finally, we should have expectations for teaching that align with expectations for students as set out in our graduation requirements. From a student’s point of view, a course is a course, regardless of whether the instructor teaches the same course or a different course the next hour, and work for that course is balanced with work for three others.

For all of these reasons, I offer a suggestion for simpler guidelines, aligned with our current course credit system. They focus on what we have in common, and could work with the "department templates" described in the EAC report. They incorporate the flexibility we like, emphasize pedagogical considerations, and give some support for senior work. I begin with the assumption that we all work hard to offer our students valuable learning experiences, while recognizing that there are resource constraints at play.
Proposal to replace the current Teaching Load Guidelines

• A full-time faculty member at Middlebury College teaches an average of 4.5 classes per year, where a class is something that a student takes for one credit toward graduation.
• Each class meets for a minimum of three hours per week. If a department wants to offer a class that meets for more than six hours per week, that department must justify the pedagogical need and explain how the extra time will not infringe on students’ ability to do the work required by their other classes.
• FYSE sections are capped at 15, CW courses and introductory language classes at 16, and senior seminars at 9. Other courses may be capped at 24.
• Each department is charged with staffing its courses in accordance with its stated learning objectives. The department chair will prepare an annual teaching plan, which will include consideration of pedagogy and equity within the department, for review by the EAC.
• An individual faculty member who teaches (solo) more than 400 students over four years is entitled to a course release; in this case the department chair must present a pedagogical justification for the situation.
• A faculty member advising senior theses may bank those theses and get a course release after advising nine.
• If the EAC and the Dean of the Faculty observe that some department is (a) not doing its fair share, (b) not treating its own members fairly, or (c) making choices that have negative consequences for the institution at large, then the EAC and the Dean may impose corrective measures on that department and that department alone.