**Grade Data 2024 (a happy but long email)**

Greetings, my Data-Loving Administrators,

As promised at the beginning of the summer, I have crunched my students' grade data and compared/contrasted/crunched-some-more with their actual AP Exam scores. The overall purpose was threefold—which is an expansion upon what I did last year for analysis, which was just the first two of these goals:

1. Assess whether the students were succeeding under my current model of gradebook—because grades correlating to exam performance doesn't do much good if all the students are doing horribly.
2. Assess whether practice AP Exams or the comprehensive class grade (with its formative assignments, late penalties, etc.) is the better predictor of AP scores.
3. New goal: assess how well various grade categories in my class—labs, quizzes, and homework—correlate to AP Exam performance.

The entire raw data file is attached in spreadsheet form. One note: the "Overall grade" is a snapshot of what the students' grades were *at the time of the AP Exams in mid-May*, and may thus be a few decimal points off of what their final end-of-year grades were—or more, if the student, say, handed in a late Lab in early June. Nonetheless, what we've got here is quite workable.

If you'd like the shortest possible summary of how it all turned out, here you go:



If that's not technical enough, here is a more detailed analysis of major points:

Regarding goal #1, student performance:

* Between all classes, there were 82 A's, 48 B's, 10 C's, 0 D's and 0 F's.
* On the AP Exams, there were **22** 5's—a personal record since I came to this school, and that's despite having an AT contract two of those years!
* Not on the spreadsheet because it's anonymized, but I did keep a count: girls slightly outperformed boys on the exams, particularly in AP Physics 1, where my overall pass rate was 56.7%, but girls passed at 59.3%. Girls were also underrepresented among the 1's (which is where you *want* to be underrepresented) in all classes.
* Also not on the spreadsheet yet: I believe (but would have to confirm, because ethnicity isn't always 100% obvious) that 100% of my Latino/Hispanic students passed the exam, and that (within margin of error) black and Asian students' performances were generally in line with that of white students. That said, we still suffer from small sample size; please tell counselors to encourage black and brown students to take AP Physics! When they get here, they're by and large succeeding, but it's hard to help students who aren't in the class to begin with. We even had one student this year (Dominic Wallington) earn a full scholarship to Lafayette College in an engineering program. There's room to grow, here!
* My AP Physics 1 pass rate grew 2.4%, which outpaced the world (1.8%) and the state of VA (1.2%). Also, the overall AP1 pass rate of 56.7% is better than the world (47.3%) and VA (45.7%) and sets a personal record for my AP1 classes. (Of note: the exam is getting a rewrite next year, with College Board pretty much apologizing to the world for the test being so hard to pass. We'll see if rates jump everywhere next year.)
* The much bigger news, though, belongs to AP Physics C, where our already great scores rose 9.3%--almost 88% of our APC students passed the exam, with 57% earning a 4 or 5! Not on the spreadsheet, but also of note: 5 out of 6 juniors in AP Physics C passed the exam. So they were right on pace with their 12th-grade peers.
* I believe this might be the largest number of Westfield AP Physics C students *ever* to earn a 5 (16 of them!), but to confirm that I'd need to see data pre-2017.
* AP Physics 2 scores dropped overall, but the sample size is small, so the percentage is misleading. Basically, only one student in AP2 earned a 1 on the exam, but 7 students (out of 17) earned a 2. That's abnormally high, but a mere difference of 2 students off last year's numbers. Variation is somewhat expected when there's only one section and it's 17 students large.

So that's pretty spiffy! But how about goal #2: comparing predictors?

* Last year, both Practice AP Exams (PEs) and Overall Grade (OG) were good predictors of students' AP Scores, but OG were just a little bit better—suggesting the formative grades, homework, extra credit, etc. were all fine and actually help the overall correlation, counter to SBG claims.
* This year, both PEs and OGs are once again good predictors. PEs win on one count: they were able to exactly match students' exam scores more often than OGs. However, OGs were more equally likely to overestimate or underestimate student scores across classes, whereas PEs *averaged out* to a better number because they tended to overestimate AP2 scores but underestimate APC scores—though it must be noted the PE predictions for AP1 were pretty spot on this year. If you're going to give PEs the overall win, in other words, it's partly a case of two wrongs accidentally making a right.
* On the other hand, OGs were better at predicting *pass rates* than PEs this year. When OG's predictions missed, it tended to be between 1s and 2s, for example, and not between 2s and 3s. To give you an idea of how this went:
	+ AP1 PE Prediction: **50.0%** pass (in other words, 50.0% of my AP1 students passed their Practice Exams with a 3 or higher). AP1 OG Prediction (i.e. "if you had a grade of \_\_\_%, you should get a \_\_\_ on the exam"): **56.7%** predicted pass rate. Actual AP1 pass rate:**56.7%** exact.
	+ AP2 PE Prediction: **64.7%** pass. AP2 OG Prediction: **52.9%** pass. Actual AP2 pass rate: **52.9%**.
	+ APC PE Prediction: **82.7%** pass. APC OG Prediction: **91.4%** pass. Actual APC pass rate: **87.9%**.
* So, PEs were slightly better at predicting exact scores, but OGs were better at predicting whether a student would pass with at least a 3. That said, the difference was so few students in each case, the honest assessment is probably that this year's score predictors are a draw.

Alright, so let's talk about those individual grade categories (goal #3). It's worth stopping a moment and explaining how each of these were scored throughout the year. Then, I'll summarize the results for each category.

* Homework was 10% of the gradebook, and consisted of two equally weighted parts. Part 1 was a daily stamp for fully attempting that day's homework through to a solution, whether or not the work was correct. Part 2 was correcting it by the day of the unit test—answers would be placed online in the interim. This is essentially a completion grade, but do note what "100%" in this category means: it means the student was consistently *trying new material*. It's crucial that in no way, shape, or form does a 100% grade here mean the student has fully learned that material. It's like giving a grade for a rough draft in an English class: it doesn't have to be perfect in order to get a perfect score; it just has to be a *perfect rough draft*. That's an advanced concept, but it's one worth wrapping our heads around, because it was crucial toward creating a sense of *buy-in* for groups of students who are historically underrepresented or disadvantaged. The overall correlation between homework completion and AP Exam score is weak due to 1. a low number of 1's in general (which isn't really a bad problem) and 2. a few outlier students who were able to get 4's despite not doing all their homework. It's worth noting, though, that the outliers (good and bad) are almost all white and Asian males. For girls and black and brown students, the correlation is much higher and the relationship much clearer between homework and exam scores. I plan on running that statistic separately—stay tuned.
* Quizzes were 20% of the gradebook, but only in AP Physics 1 and 2 (we didn't have quizzes in AP Physics C; that was more like a college class, where only comprehensive tests were given). Quizzes were all 10 questions long, and students got those questions right or wrong—no partial credit. The grade was out of 10, but it was curved: a score of 8, 9, or 10 questions correct earned a score of 10/10. 7 questions right earned 9/10. 5 or 6 questions correct earned an 8/10. 4, 3, and 2 or fewer questions earned 7, 6, or 5/10 respectively. So, once again, crucially, *a perfect grade could be earned without a perfect performance*. Once again, this was to increase buy-in; a student who gets 2 questions wrong and earns 100% is more likely to stick with the subject and learn than a student who gets 2 questions wrong and earns a B-minus. (Also, it lines up nicely with the fact that a 75% or so on the AP Exam earns a 5. But really, I think there's something to be learned here about equity: one of the most unfair things we do across class levels is curving AP and Honors classes, but expecting near-perfection in Regular courses to earn an A. Maybe more classes should be curved this way in general, so that we're nice to students for the first few mistakes and only drop scores later. Just a thought.) These quizzes very strongly correlate to student success on the AP Exams, across all demographic groups.
* Labs were 40% of the gradebook. In AP1 and AP2, that means labs were worth more than tests. Labs were graded on a 50-point rubric that stayed constant all year long. To get a perfect score on a lab, students *did* have to perform perfectly (according to the rubric). This score was also completely without curve. However, most common mistakes were minor 1- or 2-point errors (e.g. forgetting units on an axis), and in practice, grades lower than 40/50 were rare. There is one student who, due to personal circumstances, did not hand in his lab notebook regularly throughout the year, but was able to earn a 4 on the AP2 Exam. If you're willing to allow that single outlier, then the relationship and correlation between Lab score and Exam score are quite clear.
* One point in favor of those who like the word, "standards:" all three of these assignment types were firmly tied to course standards and practices. I wasn't giving homework, for example, asking students to identify their favorite fact about Isaac Newton; I was giving problems that were consistent with course standards. The same goes for labs and quizzes. If you want to discuss this further, I have a hypothesis that tying assignments to the standards increases correlation between assignment grade and standardized test performance, but does not necessarily increase the actual performance on such a test. Rather, grading all the formative steps along the way is what helps build high test performance in the end. It's just an idea, but it's perfectly consistent with all categories' data this year and last year, including the high Exam scores. Obviously, if anyone can show data from the six FCPS SBG high schools showing that they're outperforming my students, I'd love to see it. But considering that data wasn't made available last year even when Mr. DiBari asked for it, I'm declaring victory until anyone with results cares to show them and say otherwise.

A few closing thoughts:

* The retake system—using the practice AP exams as the retake of all the year's tests—worked well for a 2nd year in a row. Arguably, it worked even better the second time around, although I couldn't rightly say why; I did not significantly edit it from last year.
* Attendance was better this year overall, and that certainly helped the scores.
* One sadder note: Mr. Donnelly's AP Physics 1 students didn't do nearly as well this year on the AP Exams. Last year, his students did great; I'll have to speak with him about that, because honestly it caught me a bit by surprise. Not sure what's going on there.
* Still in the works: Extra Credit vs. AP Exam score correlations, and more sophisticated demographic comparisons (where available).
* The more I look at how well my students did, and the more I think about it, the more sick I get of the term, "mastery." I've spent an inordinate amount of time pondering over this, but hear me out: learning is a verb, not a noun. I put it to you that the reason why what I'm doing works is that everything I'm grading is measuring *whether a student is learning*, rather than sticking to purely summative concepts with a misplaced air of finality like "measuring the student's learning." Put another way, remember Goal #1 from above:



Even so, the correlations in my data between grade and exam performance did generally work out. So, by any metric you like, I'd say this year was a success. Onward!

Thank you for reading this very long email, and I hope the summer is treating you well. I'm happy to explain any parts of the (large, unwieldy) data set and analysis if there are any questions, or if you'd like me to go into more detail actually pointing things out in person instead of just writing about them. Alternately, if you just love charts, check out the "Worksheet" tab on the attached spreadsheet. Sorry that they're a little messy, and the colors for AP1, 2, and C aren't consistent between the charts—thank goodness my job isn't being an Excel expert. There's a lot of pretty data there nonetheless.

Wishing the best,

Will Keay