Grade-Based Performance Prediction at the University of Michigan

Benjamin P. Koester, William Murdock, James Fogel, and Timothy A. McKay

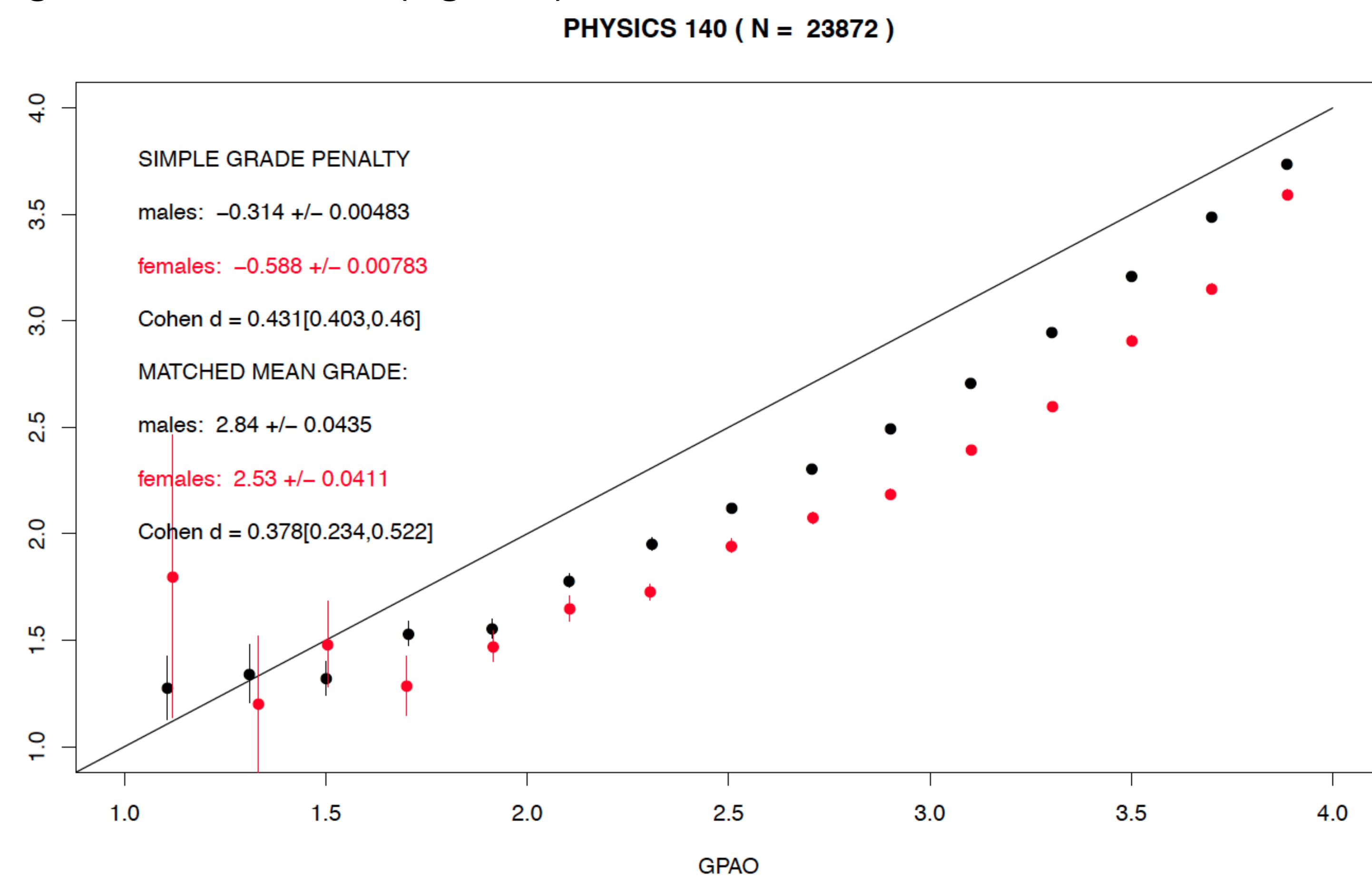*Department of Physics, University of Michigan*

Abstract

Many of the goals of learning analytics are rooted in the desire to predict student outcomes. These predictions can be used as inputs to student advising, the identification of outlier groups of students, or as a baseline to evaluate the effect of pedagogical "treatments" employed by instructors. Grades are one metric of outcome that is abundant and meticulously recorded. They contain rich information about performance and learning in a course, but also depend on the subject, term, instructor, and a student's peers. Here, we examine the predictive power of grades and show how accounting for the sources of variation underlying their distribution can be leveraged to greatly improve prediction.

Predicting Outcomes as Measured by Grades

Brute force linear regression models (not shown) that relate grades to various potential predictors (GPA, gender, ethnicity, total credits, college, and ACT/SAT composite) provide one window into the power of different predictors. They provide strong evidence that grade-point average in other classes (GPAO) is the strongest predictor of grades in future classes. This has been leveraged in studies of male and female outcomes in intro STEM classes (e.g. Physics 140). GPAO captures much but not all of the variation in grade among students: a large gap in performance between genders is observed (Figure 1).

Figure 1. GPA in other courses (GPAO) used to predict course grades. Using GPAO as a predictor, a 'grade penalty' was identified that varies with GPAO: mean grades for male and female are plotted in bins of GPAO.

- *SIMPLE GRADE PENALTY*: the average male and average female difference between their actual grade and expected grade (GPAO).
- *MATCHED MEAN GRADE*: Is the gender gap really due to gender? We match (Hansen, 2009) males and females on GPAO, ACT composite, total credits, college, and ethnicity and compare the mean grades of these matched samples and find that the gap persists.

Grades at the University of Michigan

Students' GPAs are determined by grades awarded by many different departments whose content, course structure, and grading policies differ. The mean grades for the top 10 courses by enrollment are shown for many departments, ordered by enrollment-weighted mean, in Figure 2. In addition to strong variations in mean grade between departments, some show clear grading patterns that depend on course level.
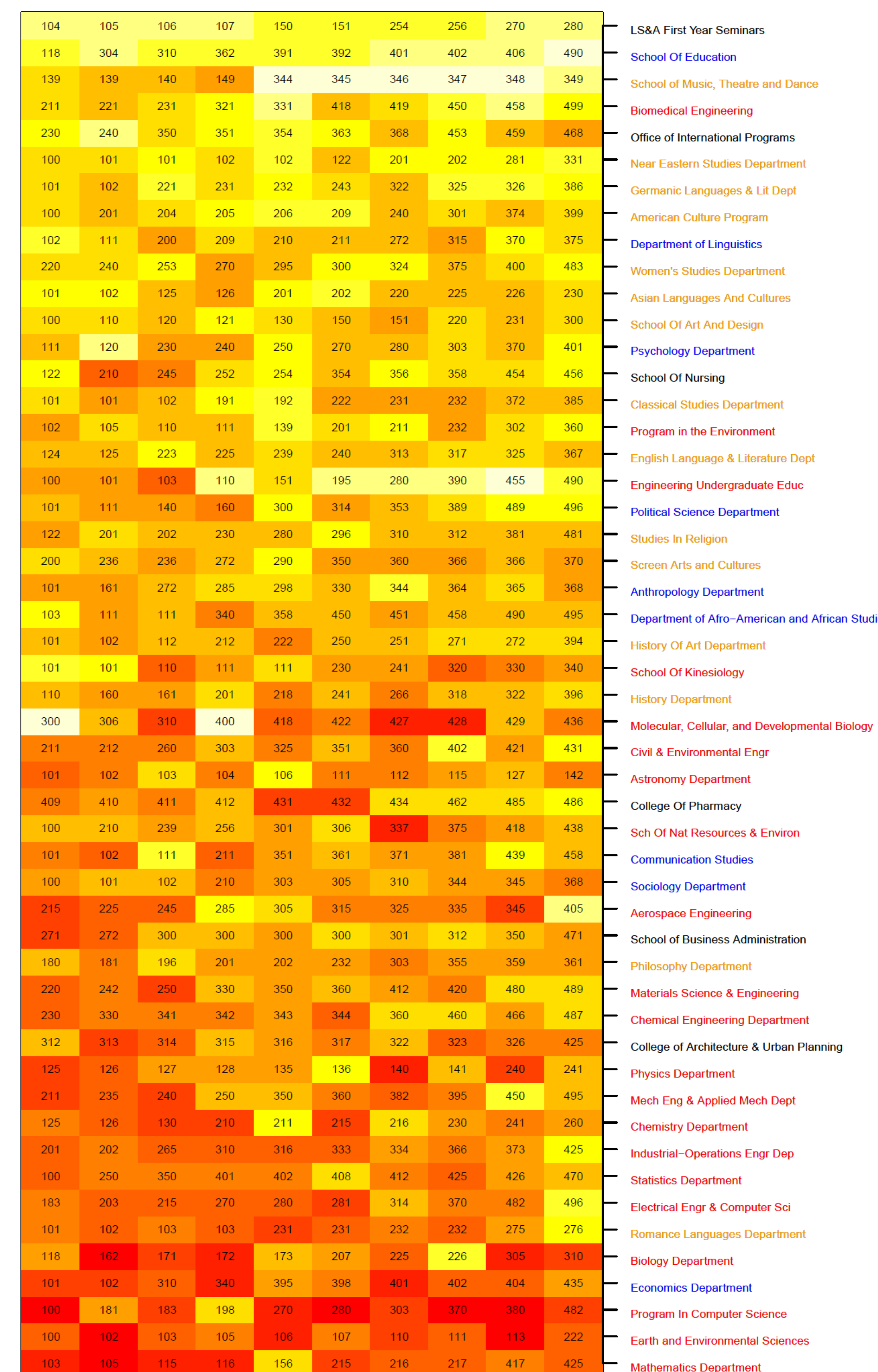

Figure 2. Mean grades in courses at Michigan since 2005. Color shading ranges from red (mean grade = 2.65) to light yellow/white (mean grade = 3.85). The top 10 courses by enrollment (left to right) are shown for departments with at least 10 courses with enrollments typically > 50. Departments are ordered top-to-bottom by enrollment-weighted mean grade. Department names are color-coded by division: science/engineering (red), social science (orange), humanities (blue), other (black).

Grade based performance prediction: GPA and beyond

GPA provides our baseline grade predictor. It is a credit hour weighted-mean of grades that is agnostic to subject, student performance, term, etc.

GPAR or grade points above replacement (e.g. Caulkins et al., 1996) provides a first refinement to performance prediction. To calculate GPAR, we form the credit weighted average of the difference between student grade and course mean grade. GPAR values center on zero, with positive scores for above average students. GPAR accounts for variations in grading practice across departments, but does not account for variation in the academic strength of student cohorts in each class.

SFE (Murdock, Fogel, et al. *in prep*) This model captures the course/term variability as well as the strength of the peer group. Each grade in a course is a linear combination of a course-term invariant student-effect and a student-invariant course-term effect:

$$Grade_{ict} = StudentFE_i + ClassFE_{ct} + \epsilon_{ict}$$

In practice, a LS solution requires manipulating a sparse $\sim 1.5 \times 10^5$ by 5×10^6 matrix. We use the sparse-matrix approximation of Arcidiacono (2012) to solve for the coefficients.

Testing performance predictors: GPA, GPAR, or SFE?

To compare these three performance predictors, we calculate them for all students based on grade data up to and including Fall 2013, then use them to predict Winter 2014 course grades. To compare, we compute the Pearson correlation coefficient ($-1 < \rho_P < 1$) between the predicted and actual grades for Winter 2014 and find that the SFE model performs substantially the better:

$$\rho^{SFE}_P = 0.734$$
$$\rho^{GPAR}_P = 0.423$$
$$\rho^{GPA}_P = 0.403$$

In individual courses, we can examine this in more detail. Each model effectively predicts a rank that we can correlate with the actual grade (or rank) in the course.
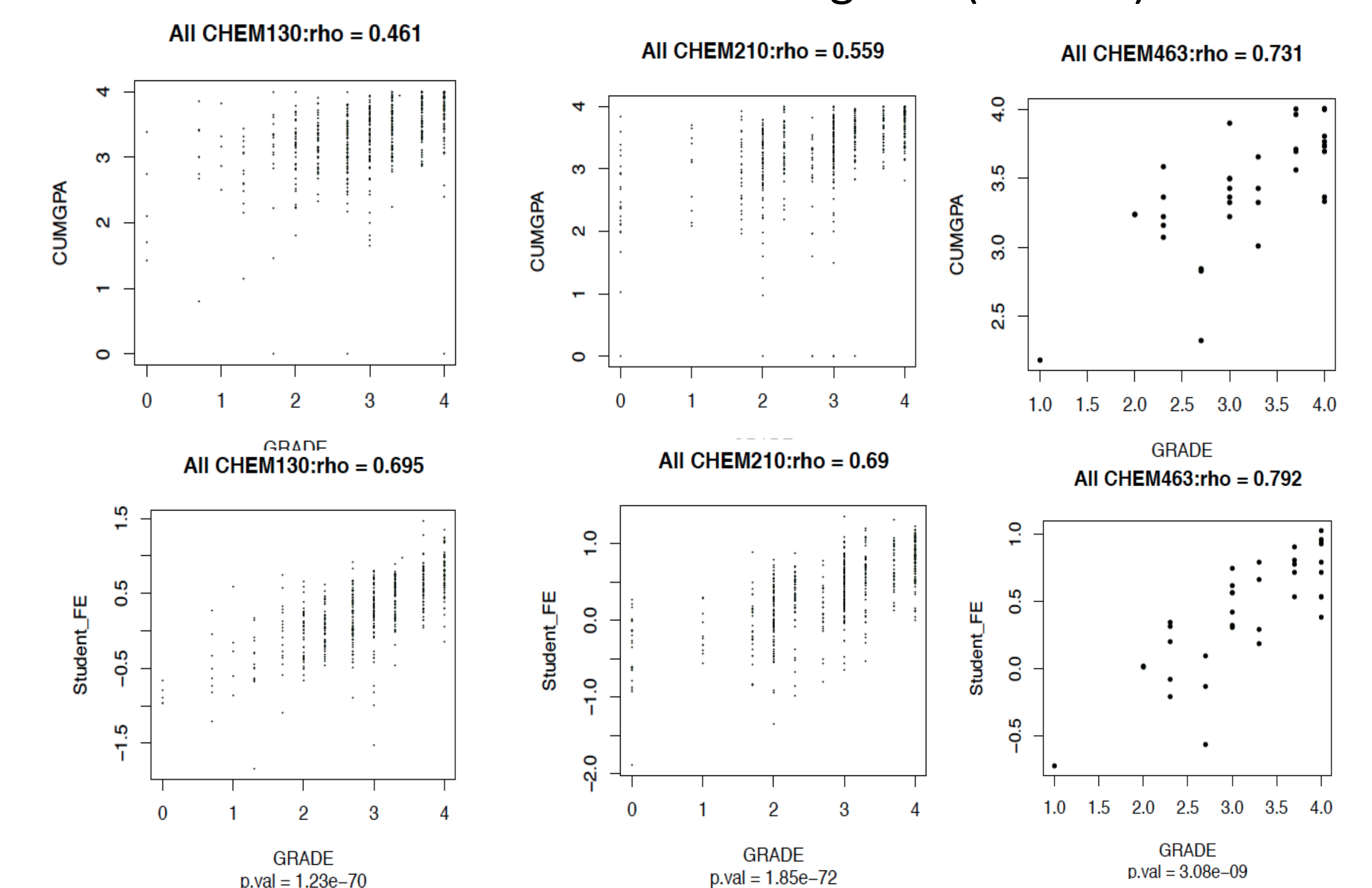

Figure 3. Single course examinations of GPA and SFE as predictors. For selected courses in Chemistry in Fall 2013, we compare GPA and SFE for their ability to predict rank, using the Spearman rank correlation between predicted and actual rank.