**Supplementary Materials**

*S1. Interpreting bik Parameters When High School Courses Are Items*

The GRM parameters of interest for course rigor analysis are *bik*, and their interpretation is analogous to typical GRM parameters (i.e., for standardized test items) under the following conditions: (1) Grade by course difficulty parameters measure grading stringency net of course-specific differences in student effort, and (2) θ is treated as time-dependent.

Interpreting *bik* parameters as measures of grading stringency net of course-specific student effort is not inconsistent with typical GRM parameter interpretation. Still, it is worth making the distinction that course rigor is operationalized as a property of an observed grade distribution, not an unobserved effort distribution. Unobserved heterogeneity in effort across courses (or items) of different difficulty levels is not a problem in GRM applications so long as one recognizes that identical *bik* parameters do not imply equality of effort*.* The GRM estimates the probability that students with a given θ would receive a grade of *k* or higher if they were to take the course. It allows one to identify dissimilar patterns of *Yij* that imply the same θ, not whether students who earned a B in AP Statistics could have earned a B+ in pre-calculus with equal effort. If grading stringency as manifest in grade distributions—conditional on θ—does not vary by course, then neither will *bik* parameters.

Unlike most educational measures, the high school GPA is an aggregation of measures from a period of several years. Time is not explicitly modeled, and *Yij* is theoretically a function of θj, item parameters, and *when students took the course*. A student who excelled in calculus at 12th grade would not necessarily have fared well in ninth grade, before they had studied advanced algebra and trigonometry. One resolution is to reinterpretθ as a time-adjusted measure of academic skill: the propensity of a student to succeed in coursework for which they have the necessary time-dependent academic skill to take. As a result, ninth graders’ θ would estimate how well they would do in calculus when prepared to take it, but it would not measure how well they would perform in calculus were they to take it in lieu of geometry in ninth grade. In the context of the high school GPA and nominal course rigor adjustment, this approach does not penalize students whose first high school math class is algebra 1 (likely putting AP calculus by 12th grade out of reach). It would presume that a student who took precalculus in 11th grade and chose not to take calculus the following year can be compared in terms of θto a student who chose to take AP calculus in 12th grade. Ultimately, while this interpretation departs from canonical GRM applications, it is consistent with conventional interpretations of the GPA construct.

*S2. Technical Details for Linear Linking of Parameters from the θ Scale to the GPA Scale*

We use a simple linear linking (Kolen & Brennan, 2004) approach to transform course difficulty parameters from a θ scale to a conventional GPA scale. More sophisticated approaches may be preferable. Our linear equation for linking to the GPA scale is , where *m* denotes the variance rescaling parameter and *z* denote the mean re-centering parameter. In this case, we have one set of parameter estimates for all courses on the θ scale, , and one set of known, policy-dictated parameters for letter grade point values in standard courses, (the vector contains the standard course grade point values plotted in Panel A of Figure 1). This allows one to use an estimator of *m* that exploits the known, fixed distance between letter grade values in standard courses on the target scale.

|  |  |
| --- | --- |
|  | (7) |

where and are two fixed points on the known GPA scale, and and are the corresponding points on the θscale for standard course *i.* In our sample, *bik* parameters are estimated imprecisely at the lower end of the distribution, so we estimate *m* using the distance between C and A grades (*bik* estimates for lower grades are also more sensitive to *a*i, as shown in Figure A2). In this case, the numerator of Equation (7) would be , or 2.0, since 4.0-2.0 = 2.0. For the eight standard courses in our sample, the average difference between and is 2.55. Therefore, 2.0/2.55 is our estimate of *m*, the factor by which the θ scale is compressed (or stretched) to match the conventional GPA scale. For our sample, Figure A1 shows the compression of points in Panel B compared to Panel A. Note that Equation (7) only depends on two fixed points, an upper and lower bound. Parameter estimates between A and C can be ignored in this estimator because the distance between A and C is equivalent to the sum of intermediate distances (i.e., (A-C) = (A-B) + (B-C)).

We estimate *z* with the mean difference in *bik* across the two scales for standard courses, after adjusting by a factor of *m*. Continuing to use *bik* for C through A as our scale anchors, the equation is:

|  |  |
| --- | --- |
|  | (8) |

where *I*={*Algebra 1, Algebra 2,…Statistics*}, *K*={*C,C+,…A*}, and there are *n* elements in *I* (*n* = 8 in our sample)and *q* elements in *K*.

Figure A1 provides a visual illustration of the scale transformation for calculus and statistics. Panel A plots , and Panel B plots . If the for AP and standard courses were the same—that is, if AP and standard courses were estimated to be equally difficult—they would fall on the main diagonal in both panels. The “1.0 Point Bonus Line” in Panel B indicates where the would lie if we had found that Advanced Placement courses in statistics and calculus were exactly one letter grade more difficult than standard courses in the same subject. If the *BONUS* quantity of interest were the difference in difficulty between an AP course and standard course in the same subject, a simple estimator for *BONUSAP\_STATS* would be the average vertical distance between the letter grades labeled “stats” and the main diagonal in Panel B.

Once all course-by-grade difficulties are on the same scale, one can estimate *BONUS* quantities of interest. Whether the correct *BONUS* quantity of interest is the difference in difficulty between an AP course and standard course in the same subject (), or the difference in difficulty between an AP course and the average standard course ()—or perhaps something else—is ultimately a policy question. Unweighted or weighted average differences in and the comparison course(s) are potential *BONUS* estimators. Possible candidates for weights could be the number of students taking the course in the population of interest or a function of the standard error for .

In our study, we estimated the unweighted average difference in between the average AP course and the average standard course, continuing to restrict *k* to letter grades C through A. In this case, the equation is:

|  |  |
| --- | --- |
|  | (9) |

where *P*={*AP Physics, AP Statistics, AP Calculus AB, AP Calculus BC*}, and there are *c* elements in *P* (*c* = 4 in our sample). Our estimate of *BONUSAP* was 0.25 letter grade points, and the analogous estimate for *BONUSHON* was 0.02 letter grade points.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE A1 |  |  |  |  |  |  |  |  |  |  |  |  |
| *Letter Grades Received by Course* | | | | | | | | | | | | |
|  | Percentage of Letter Grades Received by High School Course | | | | | | | | | | | |
|  | F | D-/D | D+ | C- | C | C+ | B- | B | B+ | A- | A | *Total (N)* |
| Algebra1 | 0.1 | 0.3 | 0.2 | 0.4 | 3.4 | 2.3 | 2.2 | 13.2 | 10.3 | 8.2 | 59.3 | 5314 |
| Algebra1-H | 0.0 | 0.0 | 0.1 | 0.1 | 1.4 | 1.4 | 1.8 | 10.5 | 9.9 | 8.3 | 66.3 | 2390 |
| Algebra2 | 0.2 | 0.6 | 0.4 | 0.8 | 3.9 | 2.9 | 3.5 | 16.8 | 11.4 | 9.2 | 50.3 | 4999 |
| Algebra2-H | 0.1 | 0.2 | 0.3 | 0.4 | 2.4 | 2.2 | 2.7 | 13.4 | 13.0 | 11.6 | 53.6 | 2923 |
| Geom | 0.0 | 0.5 | 0.2 | 0.6 | 4.1 | 2.9 | 3.2 | 16.1 | 11.3 | 8.7 | 52.4 | 5144 |
| Geom-H | 0.0 | 0.4 | 0.3 | 0.4 | 2.5 | 2.9 | 2.9 | 12.7 | 13.3 | 9.8 | 54.9 | 2822 |
| Integrat | 0.3 | 0.7 | 0.3 | 0.7 | 3.2 | 3.2 | 3.5 | 17.5 | 13.8 | 11.5 | 45.3 | 1370 |
| Integrat-H | 0.0 | 0.0 | 0.4 | 0.0 | 1.2 | 2.3 | 0.4 | 9.3 | 14.3 | 10.8 | 61.4 | 259 |
| Oth. Math | 0.0 | 0.4 | 0.2 | 0.8 | 3.3 | 2.9 | 3.7 | 14.0 | 10.3 | 7.6 | 56.7 | 485 |
| Oth. Math-H | 0.0 | 0.0 | 0.0 | 0.6 | 5.2 | 3.5 | 2.9 | 13.8 | 13.8 | 5.8 | 54.6 | 174 |
| Precalc | 0.6 | 0.9 | 0.5 | 1.0 | 4.8 | 4.3 | 4.4 | 16.4 | 12.4 | 10.3 | 44.5 | 4282 |
| Precalc-H | 0.2 | 0.5 | 0.5 | 0.6 | 2.9 | 3.1 | 3.8 | 14.4 | 14.5 | 10.7 | 48.8 | 2554 |
| Statistics | 0.3 | 0.4 | 0.1 | 0.2 | 3.5 | 2.4 | 3.3 | 14.3 | 13.4 | 9.1 | 52.9 | 911 |
| Statistics-H | 0.0 | 0.0 | 0.0 | 1.0 | 2.4 | 2.9 | 3.9 | 8.7 | 8.7 | 6.3 | 66.0 | 206 |
| Calculus | 1.1 | 1.0 | 0.9 | 0.8 | 4.2 | 3.6 | 3.6 | 14.9 | 12.8 | 10.7 | 46.5 | 1291 |
| Calculus-H | 0.7 | 1.4 | 0.5 | 0.9 | 4.7 | 2.2 | 4.3 | 13.7 | 13.0 | 7.9 | 50.5 | 554 |
| AP CalcAB | 0.7 | 1.1 | 0.5 | 0.9 | 5.8 | 4.4 | 5.0 | 15.4 | 12.5 | 10.9 | 42.8 | 2032 |
| AP CalcBC BC | 0.0 | 1.2 | 0.5 | 0.9 | 4.7 | 5.0 | 4.7 | 17.5 | 16.0 | 13.7 | 35.9 | 424 |
| AP Stats | 0.5 | 1.2 | 0.4 | 1.4 | 4.7 | 2.3 | 3.9 | 13.6 | 14.8 | 12.2 | 44.9 | 770 |
| AP Physics | 0.3 | 0.3 | 0.3 | 0.7 | 3.7 | 3.5 | 4.4 | 16.1 | 15.2 | 12.4 | 43.1 | 870 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Column %* | 0.2 | 0.6 | 0.3 | 0.6 | 3.6 | 3.0 | 3.3 | 14.7 | 12.2 | 9.7 | 51.8 | 100% |
| *Column n Total* | 91 | 219 | 130 | 251 | 1435 | 1178 | 1317 | 5846 | 4835 | 3869 | 20603 | 39774 |

|  |  |  |  |
| --- | --- | --- | --- |
| TABLE A2  *OLS Regression Prediction of Final Grade in College Calculus by GPA* | | | |
|  | GRM () | Unweighted | Weighted |
|  | (1) | (2) | (3) |
| GRM-GPA | 5.079\*\*\* |  |  |
|  | (0.209) |  |  |
|  |  |  |  |
| Unweighted GPA |  | 8.239\*\*\* |  |
|  |  | (0.397) |  |
|  |  |  |  |
| Weighted GPA |  |  | 7.807\*\*\* |
|  |  |  | (0.358) |
|  |  |  |  |
| SAT-M | 0.029\*\*\* | 0.031\*\*\* | 0.027\*\*\* |
|  | (0.002) | (0.002) | (0.002) |
|  |  |  |  |
| Parent has BA | 1.379\*\*\* | 1.279\*\*\* | 1.156\*\*\* |
|  | (0.334) | (0.333) | (0.331) |
|  |  |  |  |
| Intercept | 61.168\*\*\* | 30.976\*\*\* | 33.272\*\*\* |
|  | (1.352) | (1.744) | (1.279) |
| N | 7021 | 7021 | 7021 |
| *R2* | 0.314 | 0.302 | 0.309 |
| *Note.* All specifications include fixed effects that identify unique college-instructor-course combinations. Heteroskedasticity-robust standard errors clustered by college in parentheses.  + *p* < 0.10, \* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001 | | | |

Macintosh HD:Users:johnhansen:Dropbox:IRT_GPA:jebs:figures:theta_to_gpa.pdf

FIGURE A1. Comparison across scales of estimated *bik* parameters in calculus and statistics for standard and Advanced Placement courses. For calculus, AP Calculus AB is plotted. (A) θ scale. (B) GPA scale.

/Users/jdh488/Dropbox/IRT_GPA/measurement/figures/misc/delta/jebs_acons_v1.pdf

FIGURE A2. Comparison of *bik*parameters estimated from a graded response model in which *a* parameters are allowed to vary by course (y axis) and a graded response model in which *a* parameters are constrained to be equal for all courses (x axis).