The Case for Grades

David Eubanks

Furman University
david.eubanks@furman.edu

Course grades are free data associated with courses and information on the students who take them, and this session shows how an assessment office can make good use of this information. The session will show several ways to use grades and course registration data to identify barriers to student success and learning. Real examples will be used to show how to compute course difficulty, relative student abilities, and grading consistency using hierarchical regression. New and published work shows the links between grades and other outcomes, like learning measures, retention, teaching styles, and post-graduation success.

Course Grades as Learning Assessments

In the early days of the assessment movement, campus assessment practices were consciously separated from what went on in the classroom. This separation helped increase the credibility of the generated evidence because, as "objective" data-gathering approaches, these assessments were free from contamination by the subject they were examining. <u>Partly as a result, assessment practitioner rhetoric at the time strongly criticized grades as a valid and reliable measure of student learning.</u> (p. 19, emphasis added)

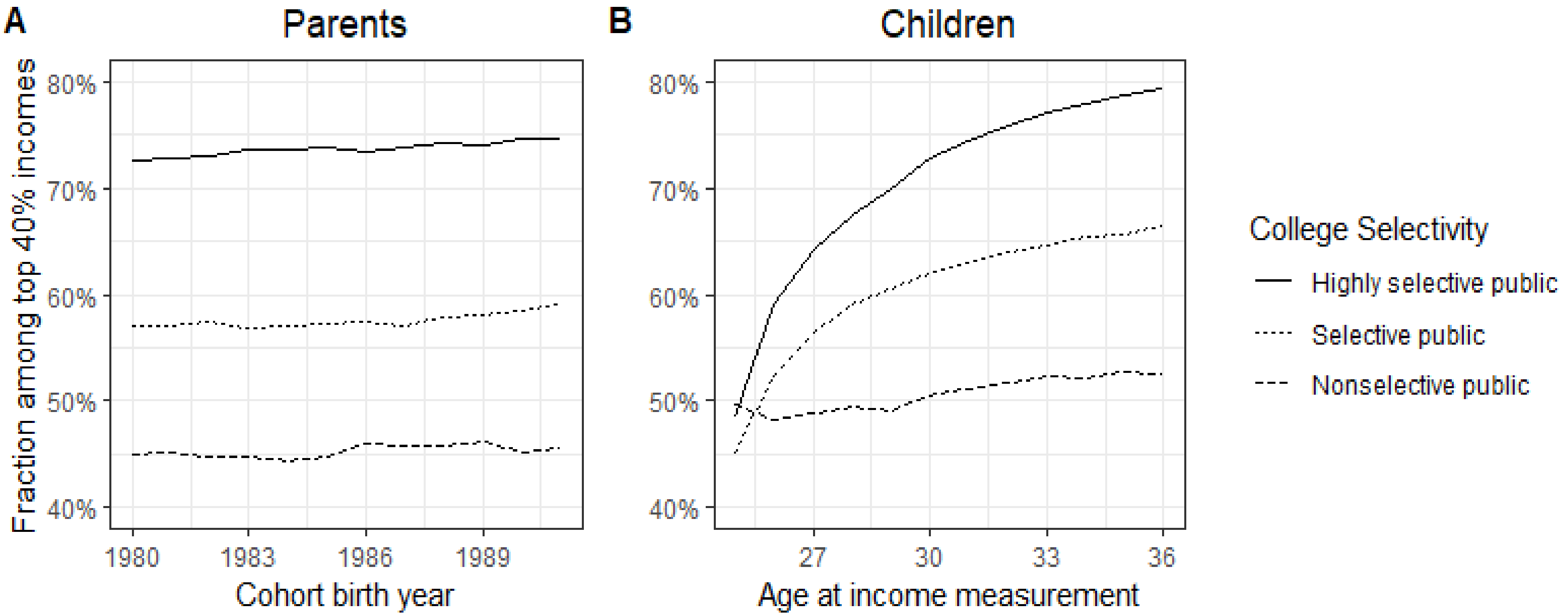
Ewell, P. T. (2009, November). Assessment, accountability, and improvement: Revisiting the tension. (Occasional Paper No. 1). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Objections to Grades

1. course grades can be rather arbitrary (just ask any student),
2. course grades usually signify the perception of an individual instructor rather than the evaluative consensus of faculty as a whole,
3. course grades are notorious for inflation ("Grades here run the gamut from A- to A+"), and
4. course grades focus on individual students and individual courses, rather than on the goals embedded in an entire degree program.

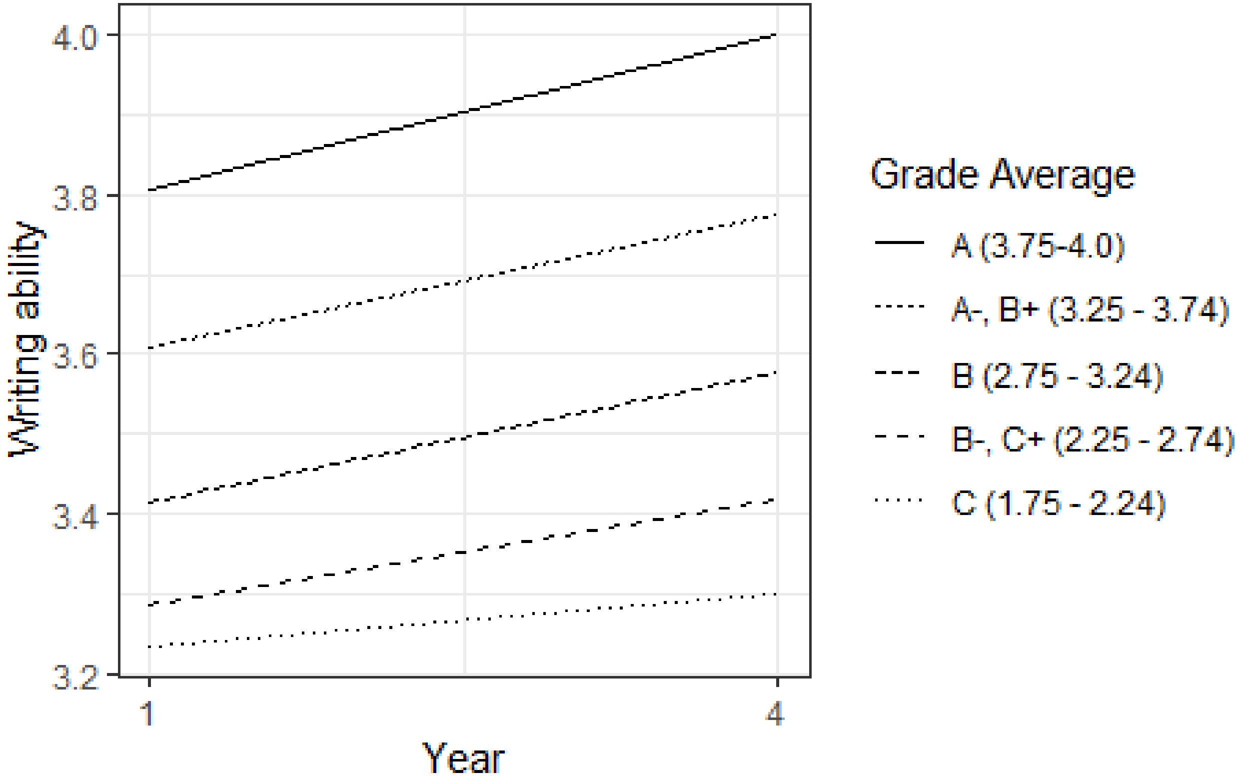
Self-study handbook for the Association of Theological Commission on Accreditation

Wealth Disparity

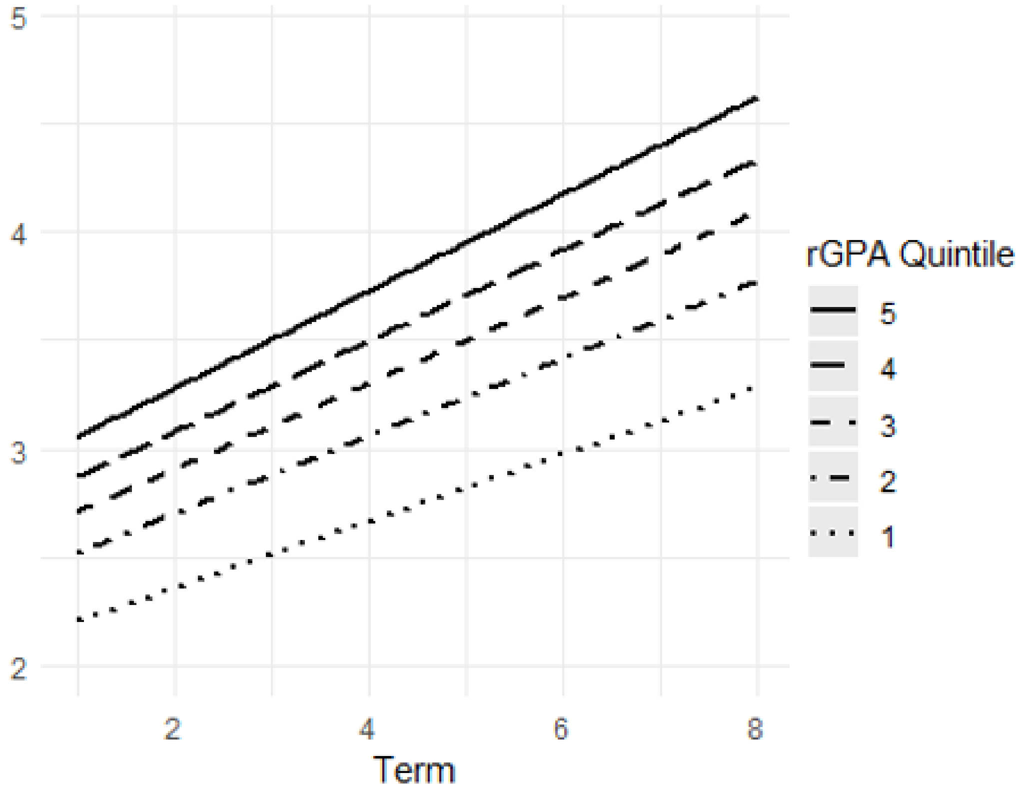


Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2017). *Mobility report cards: The role of colleges in intergenerational mobility* (No. w23618). National Bureau of Economic Research.

Writing Assessments and GPA

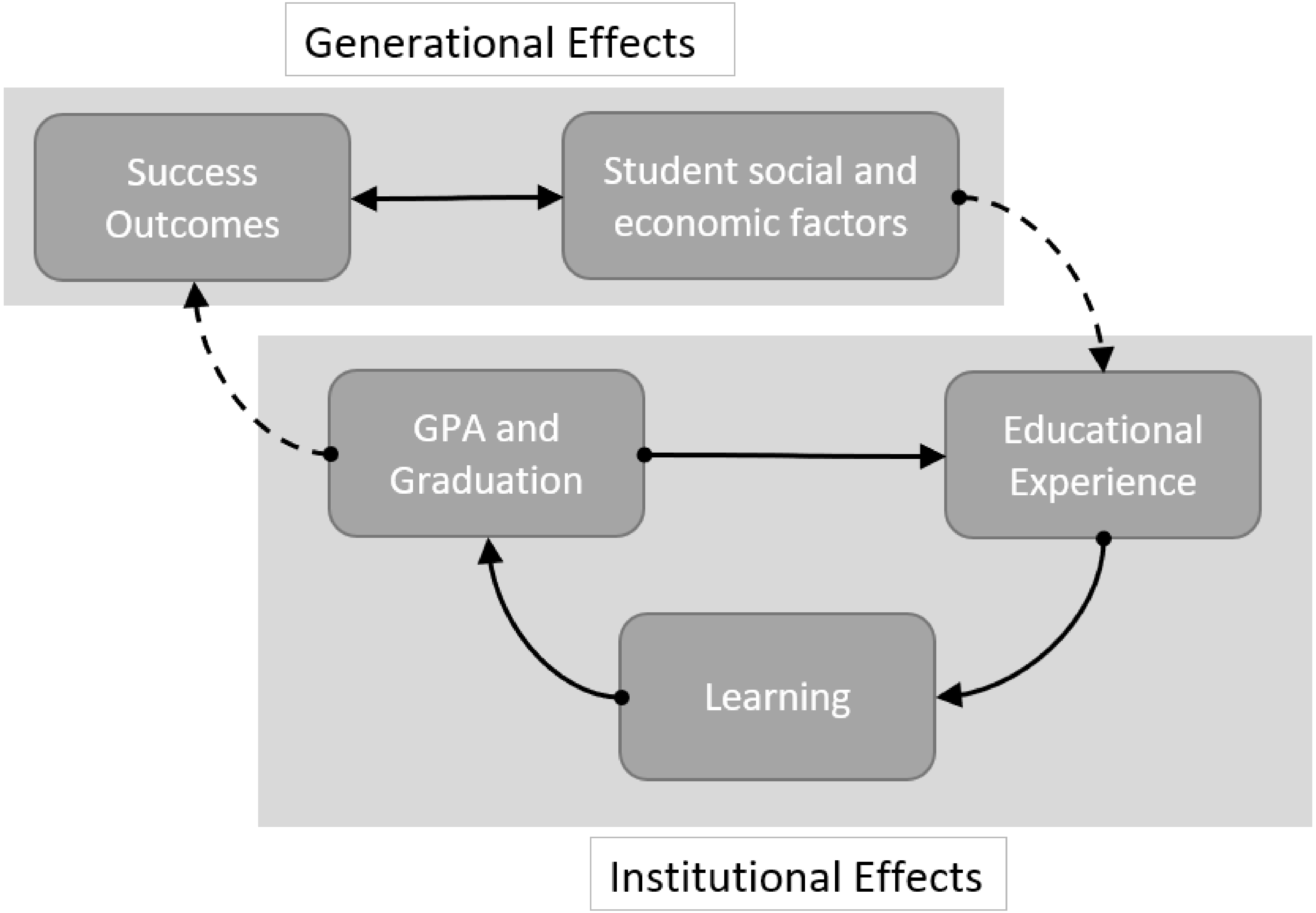


HERI National Surveys
(N = 246,000)

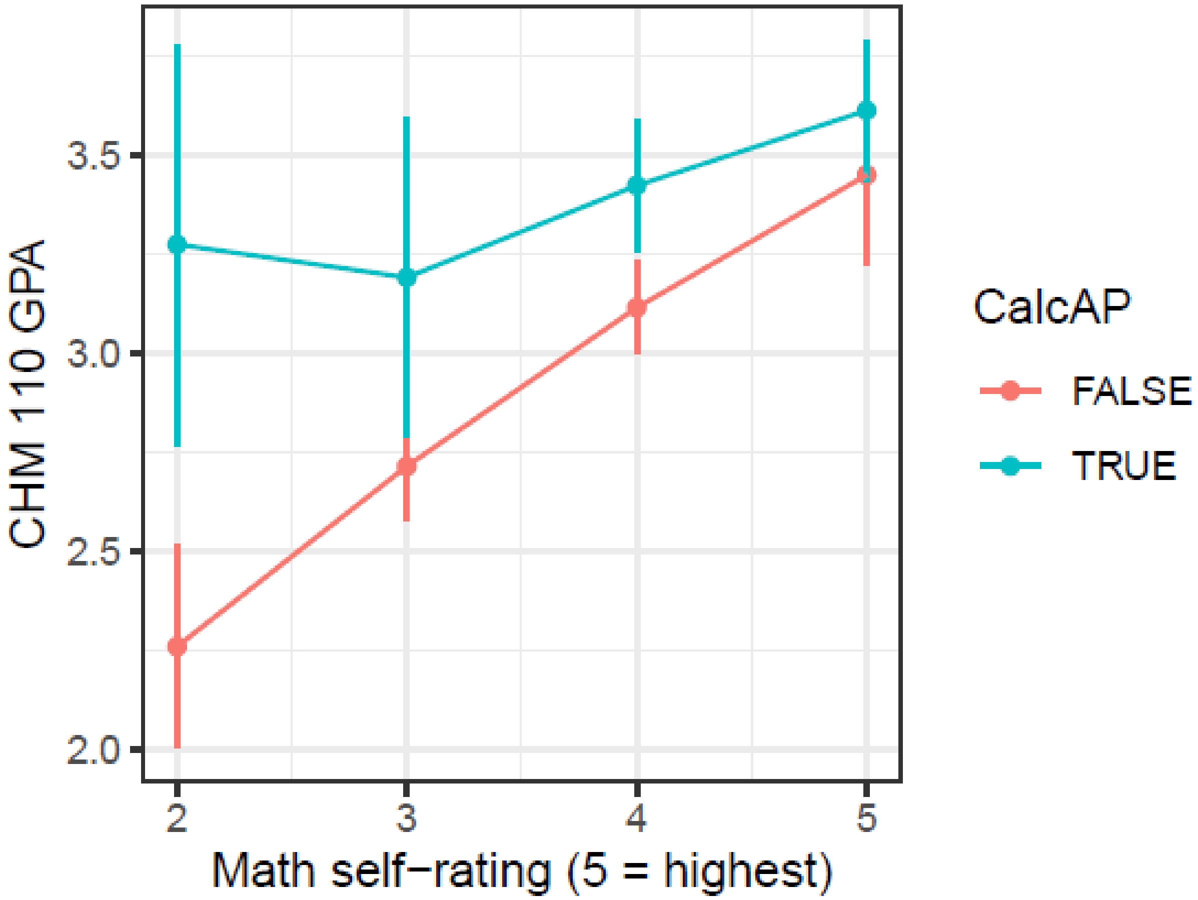


Furman University
(N = 18,000)

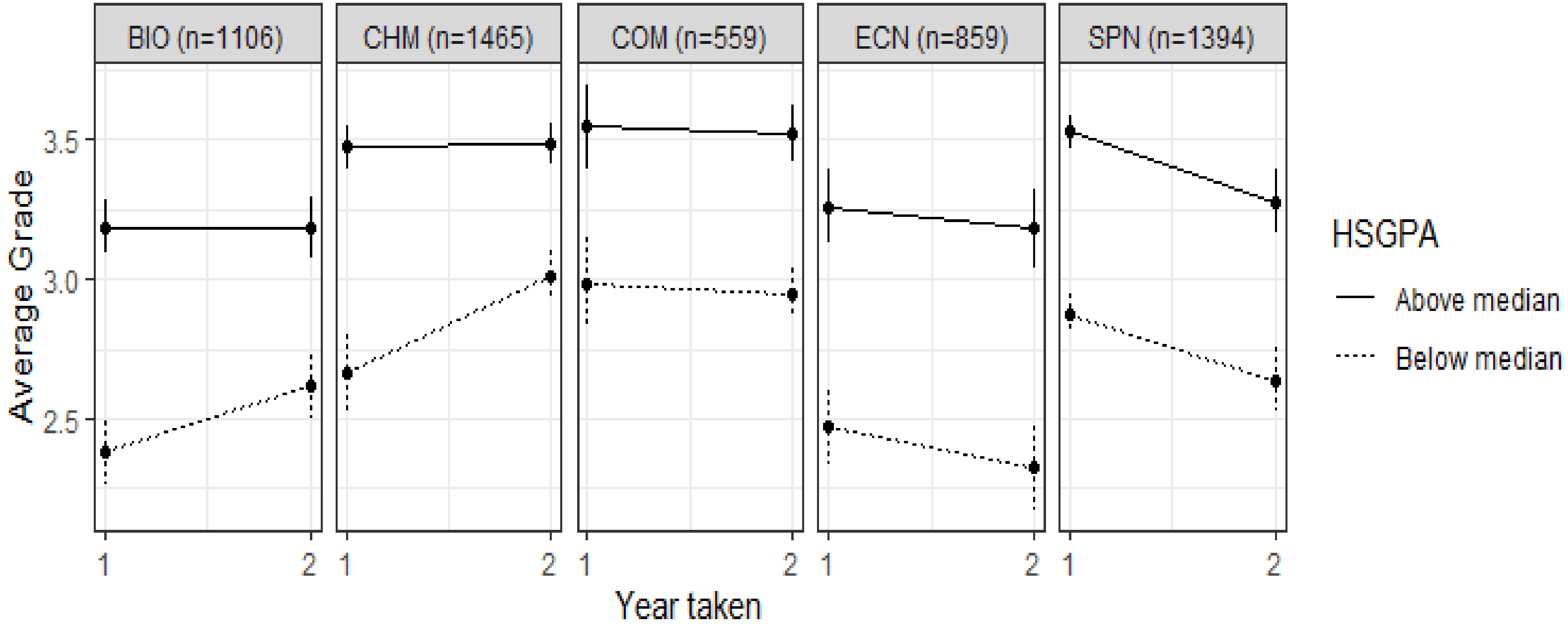
Eubanks, D., & Vanovac, S. (2020). Divergent Writer Development in College. *The Journal of Writing Analytics*, *4*, 15-53.


Generational Effects
Success Outcomes
Student social and economic factors
GPA and Graduation
Educational Experience
Learning
Institutional Effects

Math Ability and Chemistry Grades

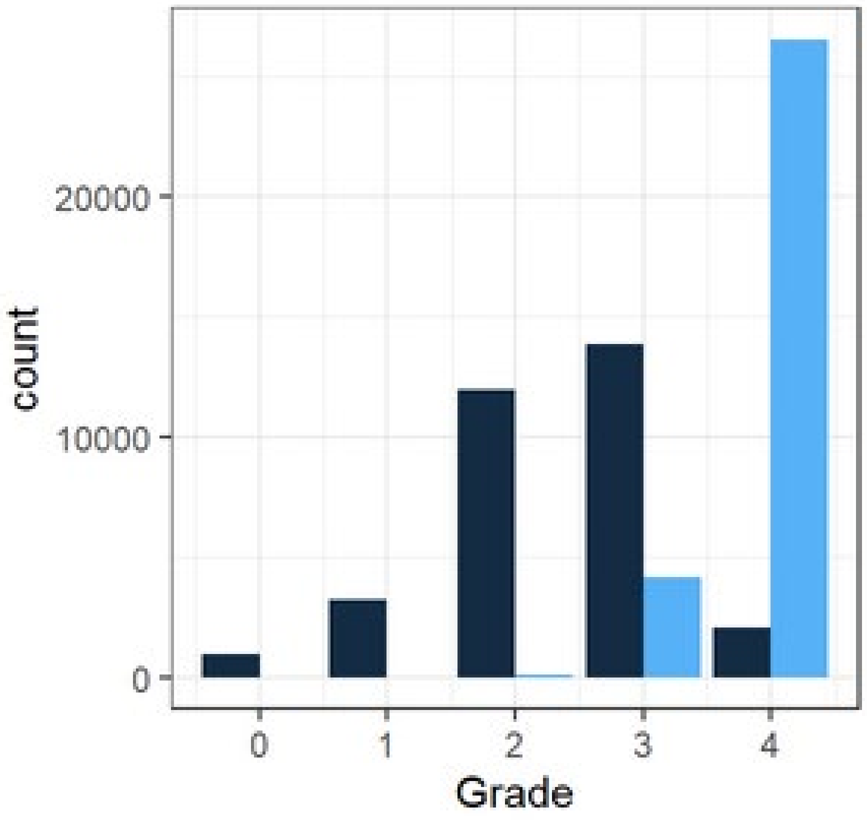


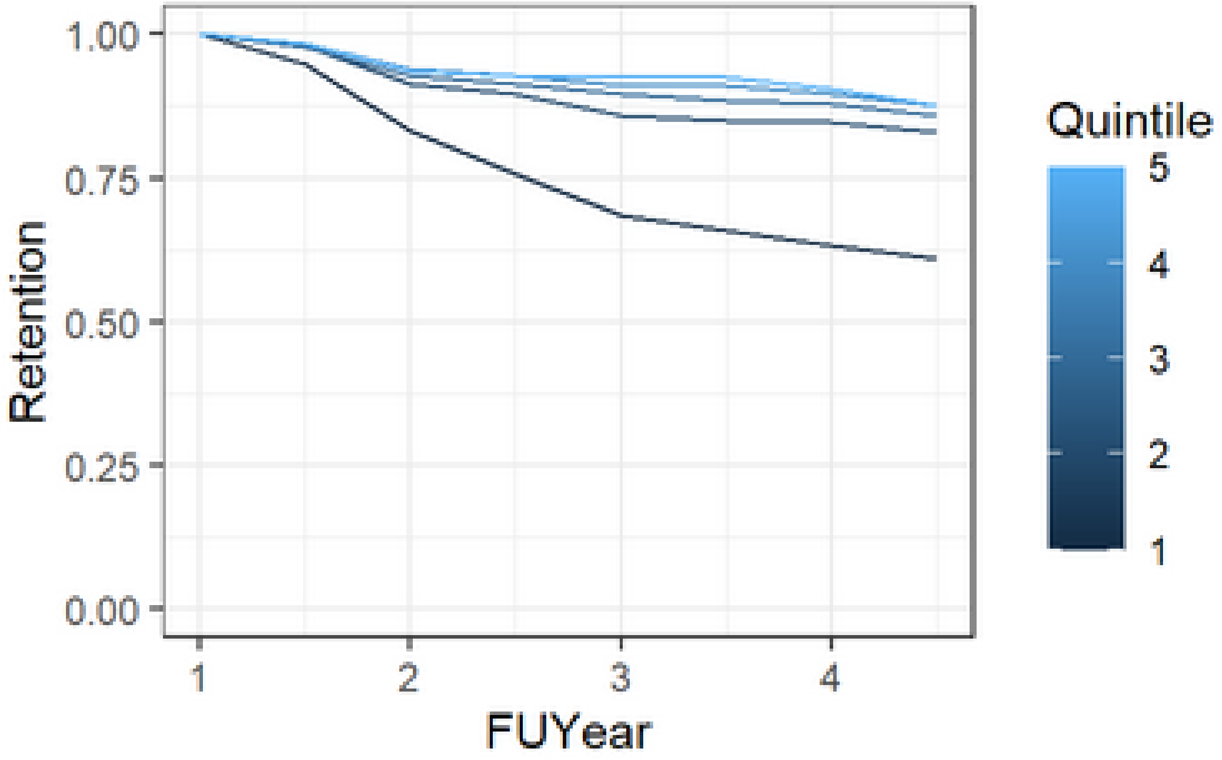
Grades and Timing



Eubanks, D. (2021) Assessing for student success. *Intersection: A Journal at the Intersection of Assessment and Learning*, 2(2).

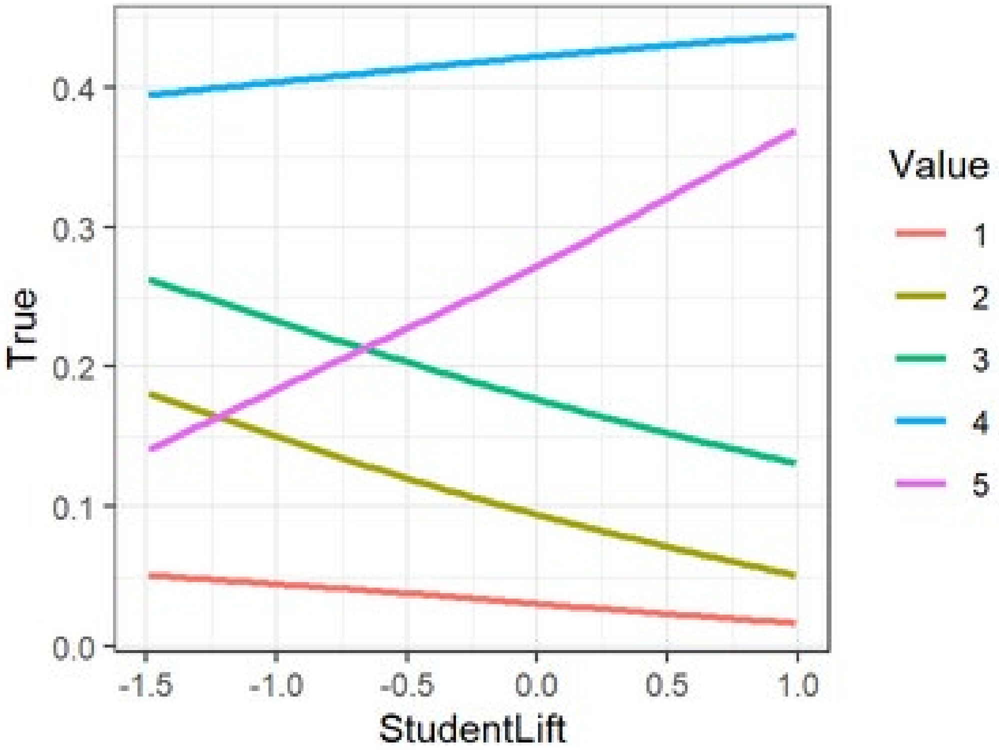
Grades and Retention



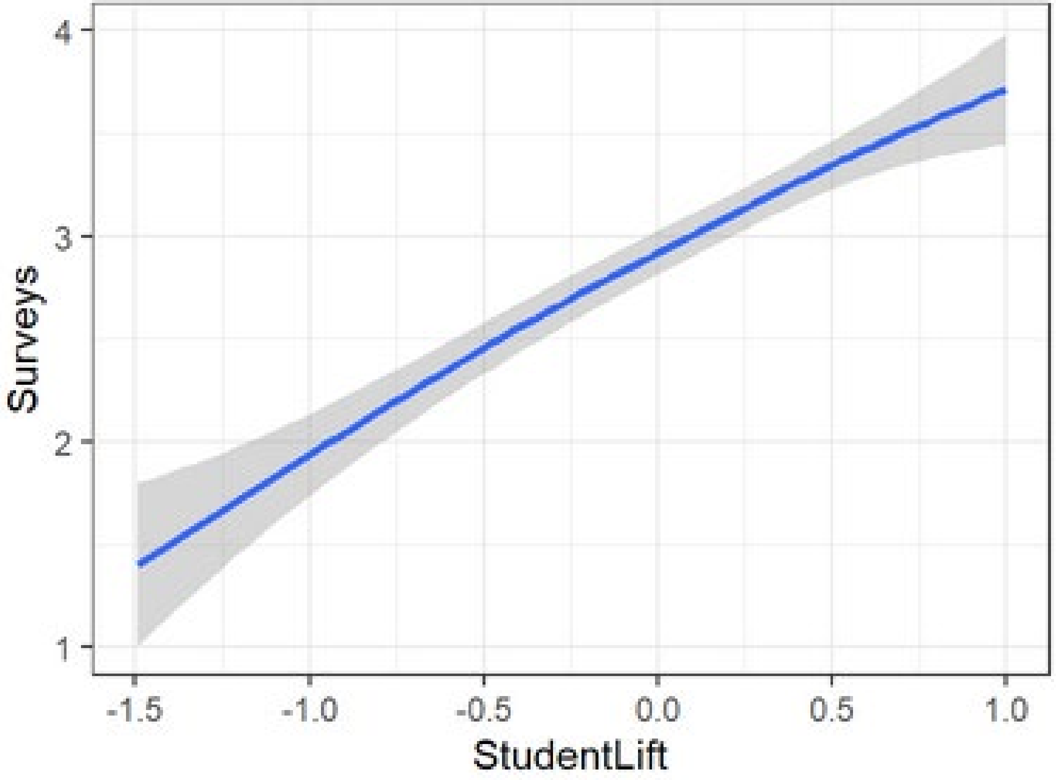


Retention rates over four years by student lift quintiles over.

Belongingness



Survey Completion



Course Difficulty

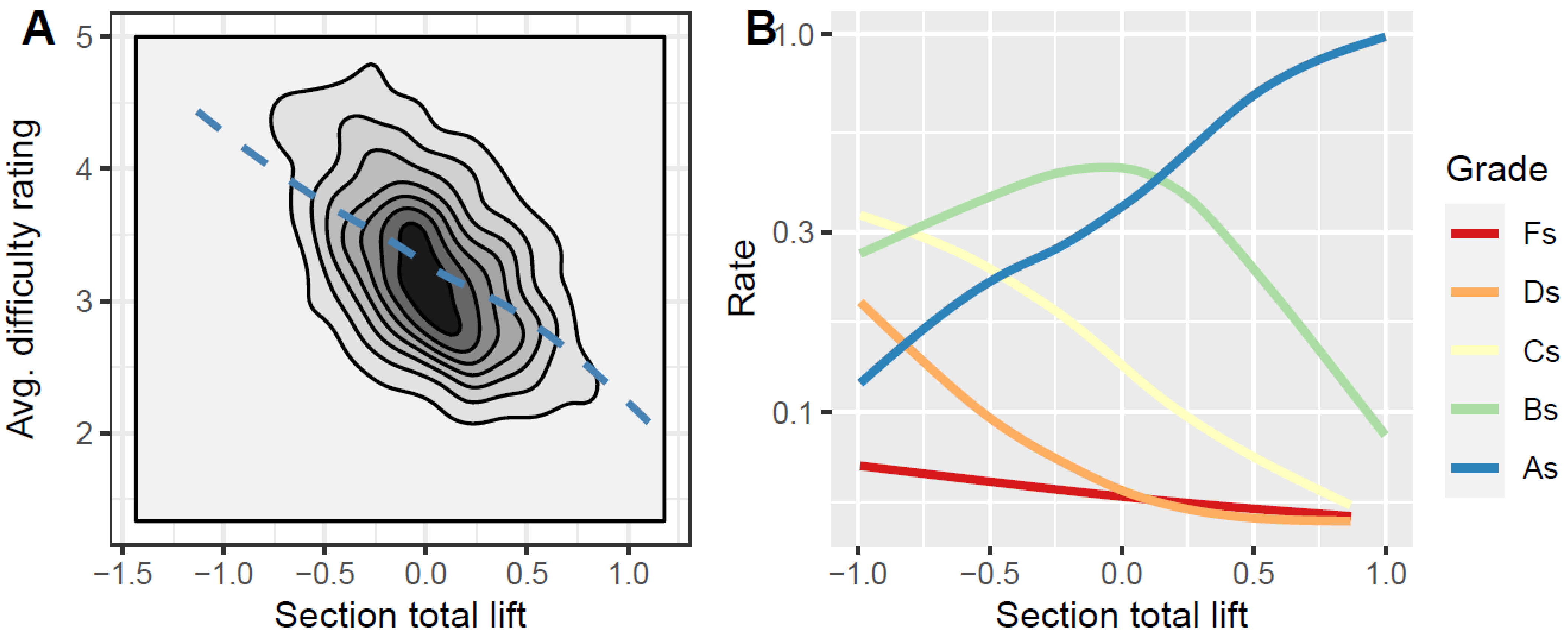
Calculating course difficulty using grades has several benefits, like creating an adjusted student GPA.

For each course section:

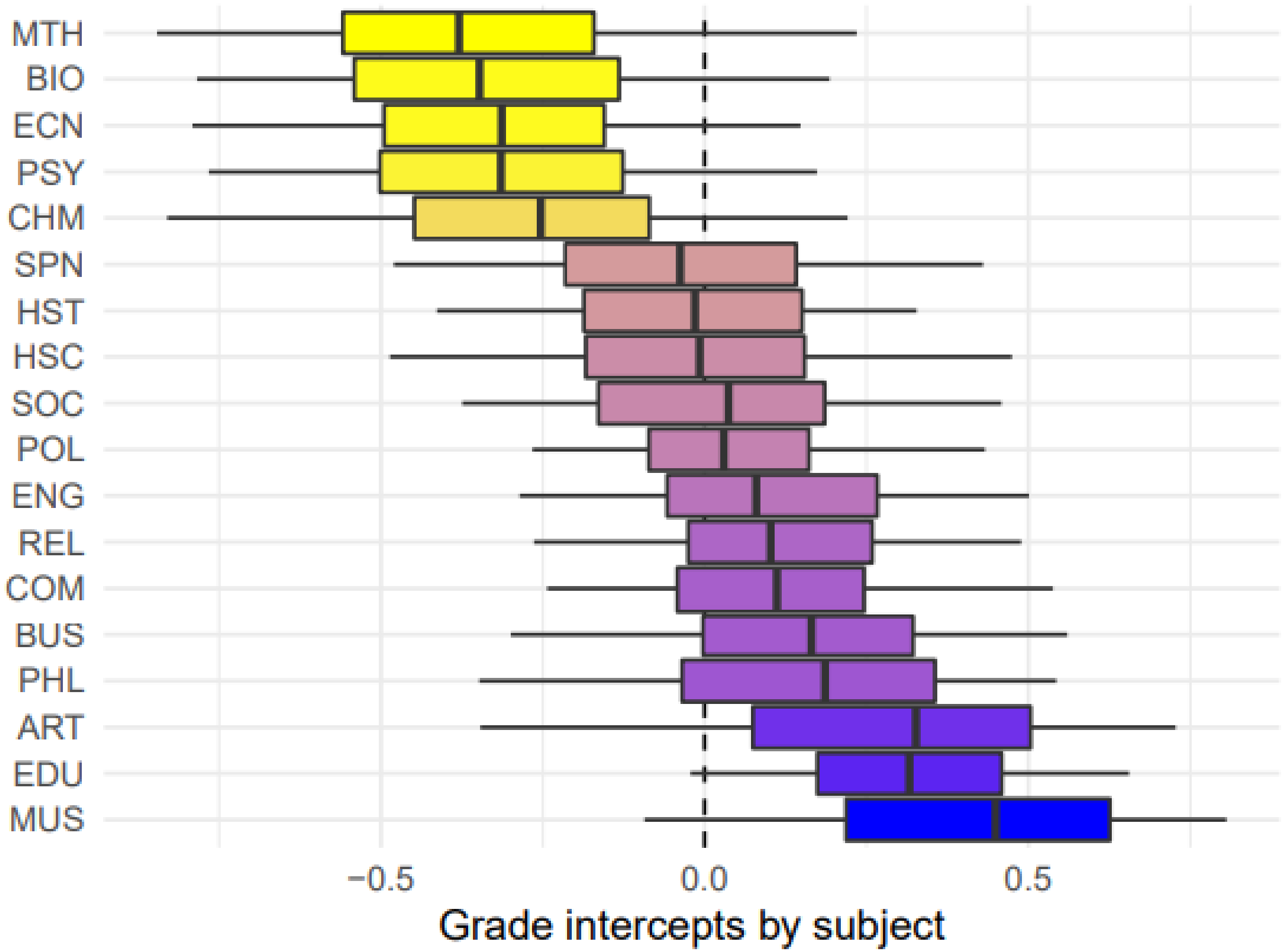
1. Find the cumulative GPA of each student in the section
2. Average these to get the expected GPA for the section
3. Calculate the actual GPA of the section from grades assigned
4. Subtract GPA – expected GPA to get “grade lift”

When lift is positive, it means it was less difficult than expected to earn grades, etc.

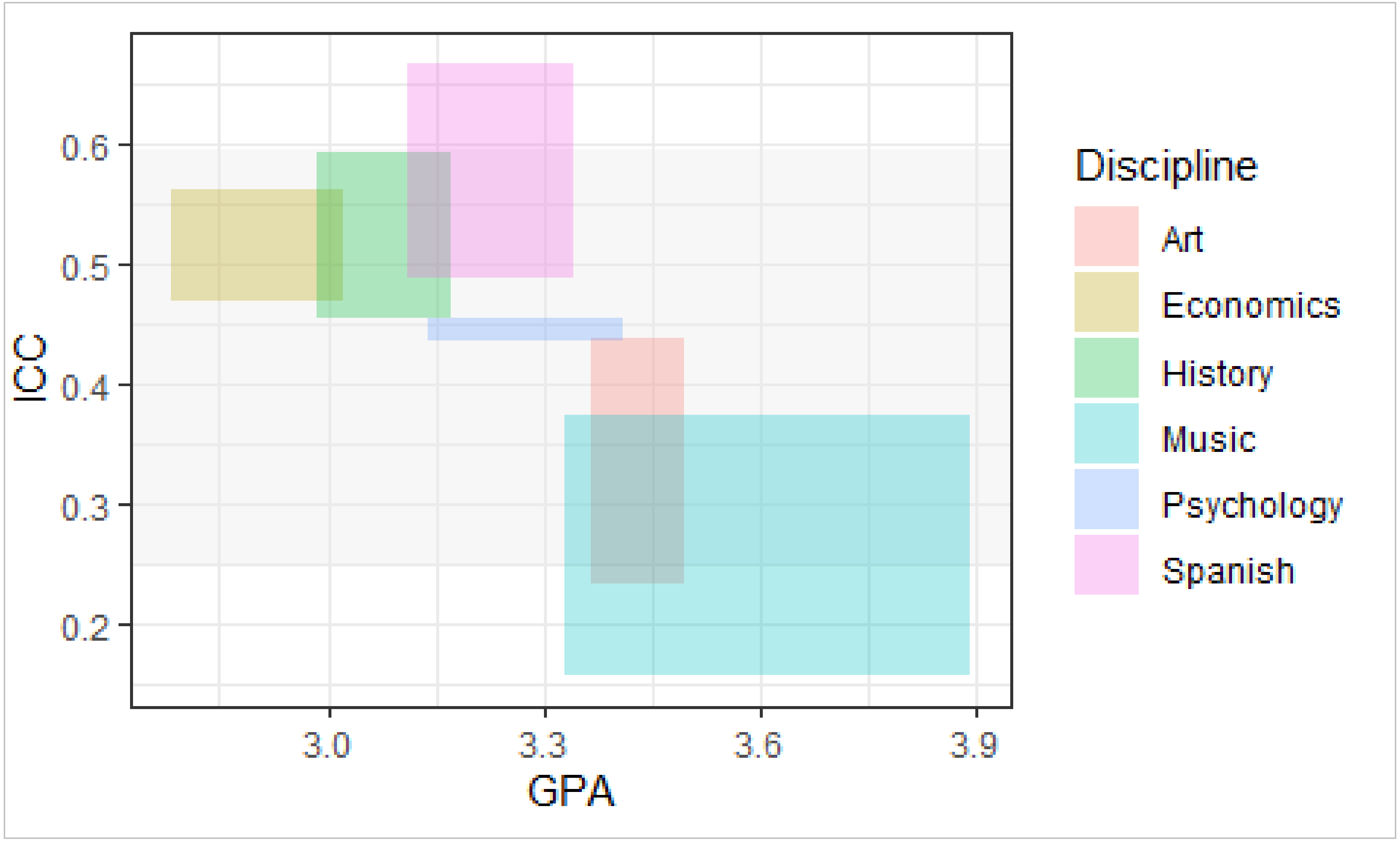
Measuring Course Difficulty



Grade Difficulty by Subject



Grade Variations and Averages



Dept. Grade Summaries

Overall Distribution

A	B	C	D	F	P	GPA	Expected
311	180	29	7	4	5	3.44	3.19

Grades by course level

As noted above, majors courses tend to have higher grades than introductory or service courses.

Level	A	B	C	D	F	P	GPA	Expected
100	78	42	5	2	0	0	3.49	3.28
200	171	118	19	1	3	4	3.42	3.16
300	27	18	4	4	0	1	3.23	3.12
400	35	2	1	0	1	0	3.69	3.25

Faculty Grade Distributions

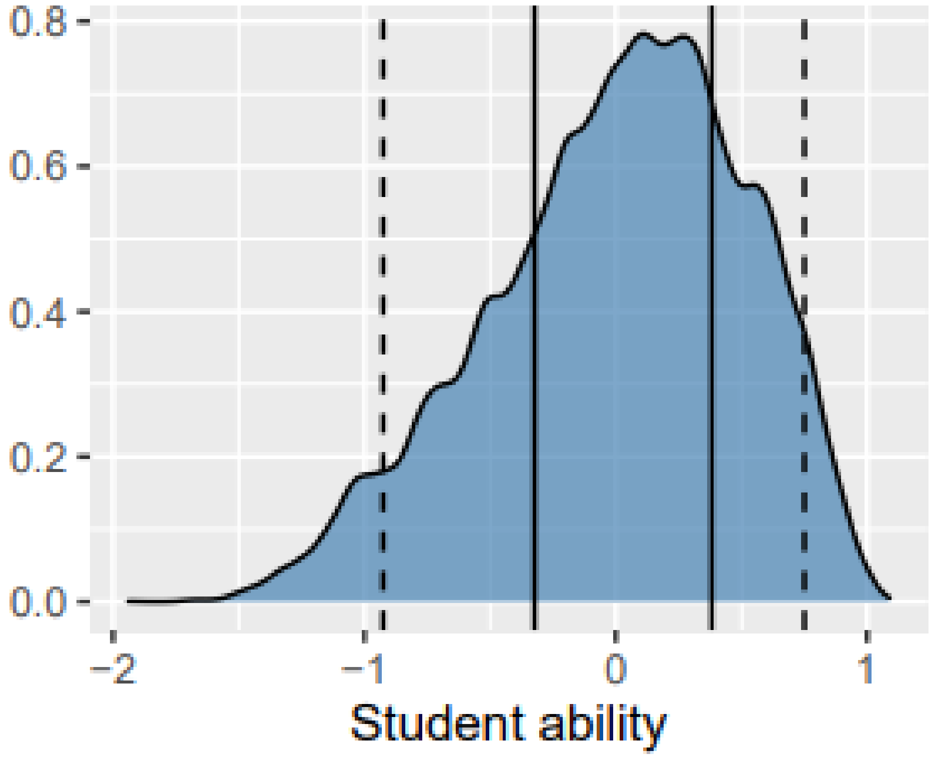
Faculty	A	B	C	D	F	P	GPA	Expected
[redacted]	1	0	0	0	0	0	4.00	3.78
	142	61	4	1	2	4	3.56	3.15
	25	15	8	0	0	0	3.32	3.28

Millet, I. (2010). Improving Grading Consistency through Grade Lift Reporting. *Practical Assessment, Research & Evaluation*, *15*(4).

Grade = Student + Section + Subject + residual

Table 2: Grade prediction linear model

	Dependent variable:
	Grade
StudentLift	1.000*** (0.995, 1.005)
SectionLift	1.000*** (0.992, 1.008)
SubjectLift	1.000*** (0.988, 1.012)
Constant	3.211*** (3.208, 3.213)
Observations	149,685
R^2	0.619
Adjusted R^2	0.619
Residual Std. Error	0.455 (df = 149681)
F Statistic	81,099.210*** (df = 3; 149681)
Note:	*p<0.1; **p<0.05; ***p<0.01

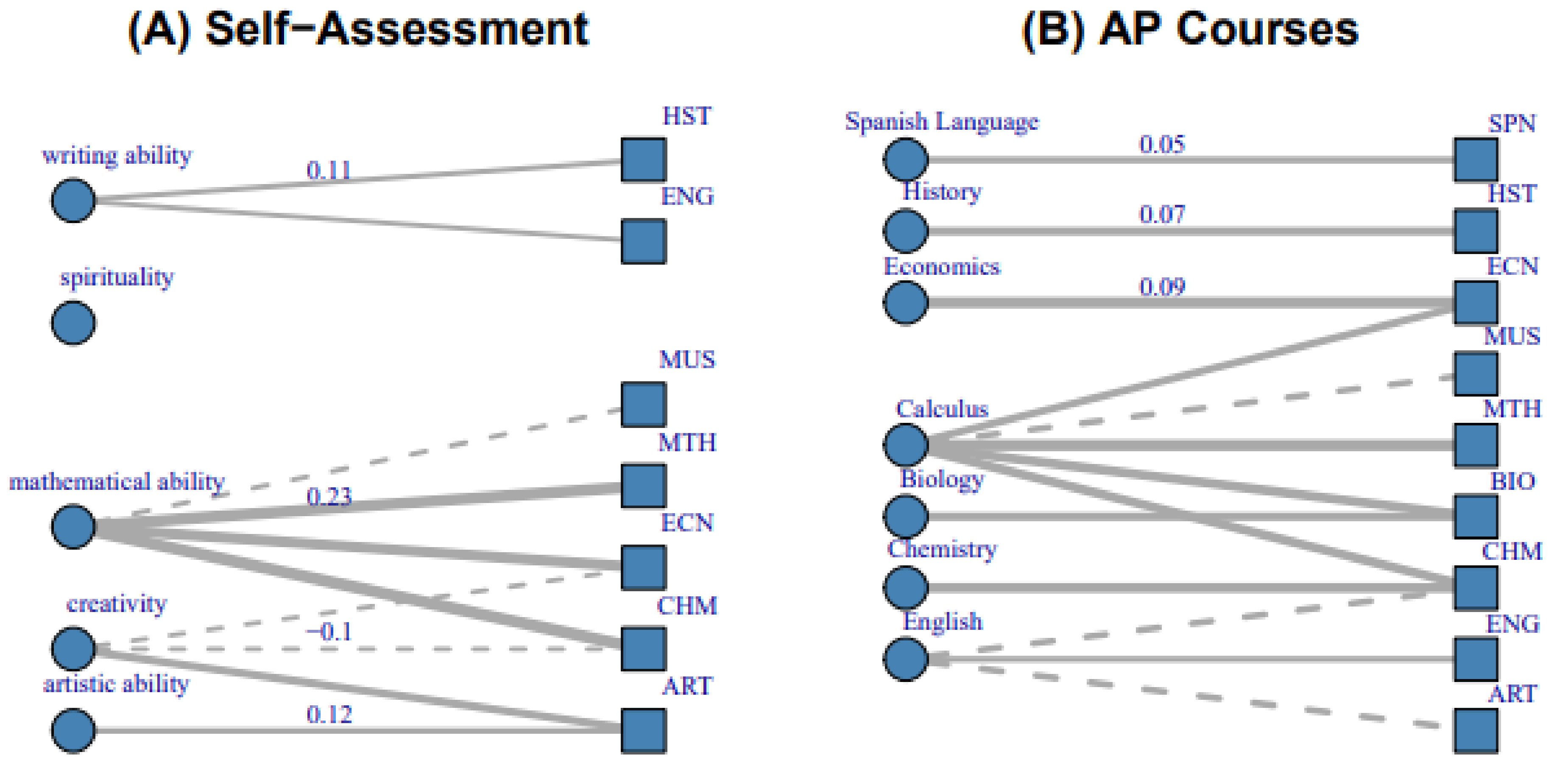


Predicting Grade-Earning Ability

Table 3: Modeling GPA or StudentLift using admissions data

	Dependent variable:	
	GPA	StudentLift
	(1)	(2)
HSGPA	0.761*** (0.692, 0.831)	0.702*** (0.640, 0.765)
AP_N	0.100*** (0.072, 0.128)	0.102*** (0.077, 0.127)
SexM	−0.094*** (−0.144, −0.043)	−0.071*** (−0.115, −0.026)
RaceOther	−0.067 (−0.209, 0.076)	−0.133** (−0.260, −0.005)
RaceWhite	0.011 (−0.117, 0.139)	−0.036 (−0.150, 0.078)
HSQuality	0.546*** (0.405, 0.688)	0.473*** (0.347, 0.600)
FirstGenFlag	−0.138*** (−0.220, −0.057)	−0.137*** (−0.210, −0.065)
TestOptional	−0.162*** (−0.220, −0.104)	−0.157*** (−0.208, −0.105)
I(log(AidAwarded + 1))	0.050** (0.006, 0.094)	0.038* (−0.002, 0.077)
Constant	−0.510* (−1.038, 0.019)	−3.240*** (−3.711, −2.768)
Observations	1,049	1,049
R^2	0.513	0.537
Adjusted R^2	0.508	0.533
Residual Std. Error (df = 1039)	0.387	0.345
F Statistic (df = 9; 1039)	121.414***	133.981***
Note:		*p<0.1; **p<0.05; ***p<0.01

Lift Residuals and Subjects



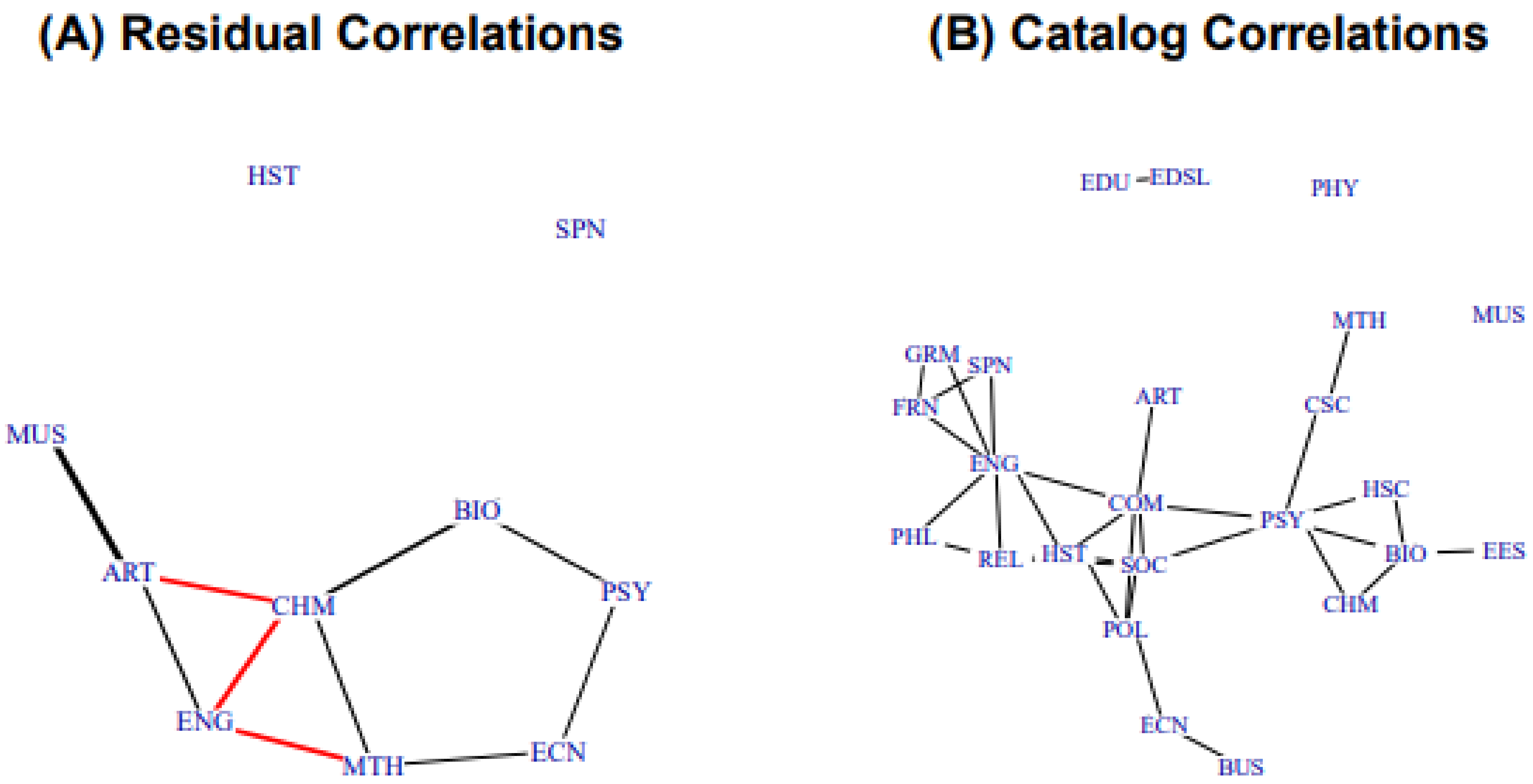


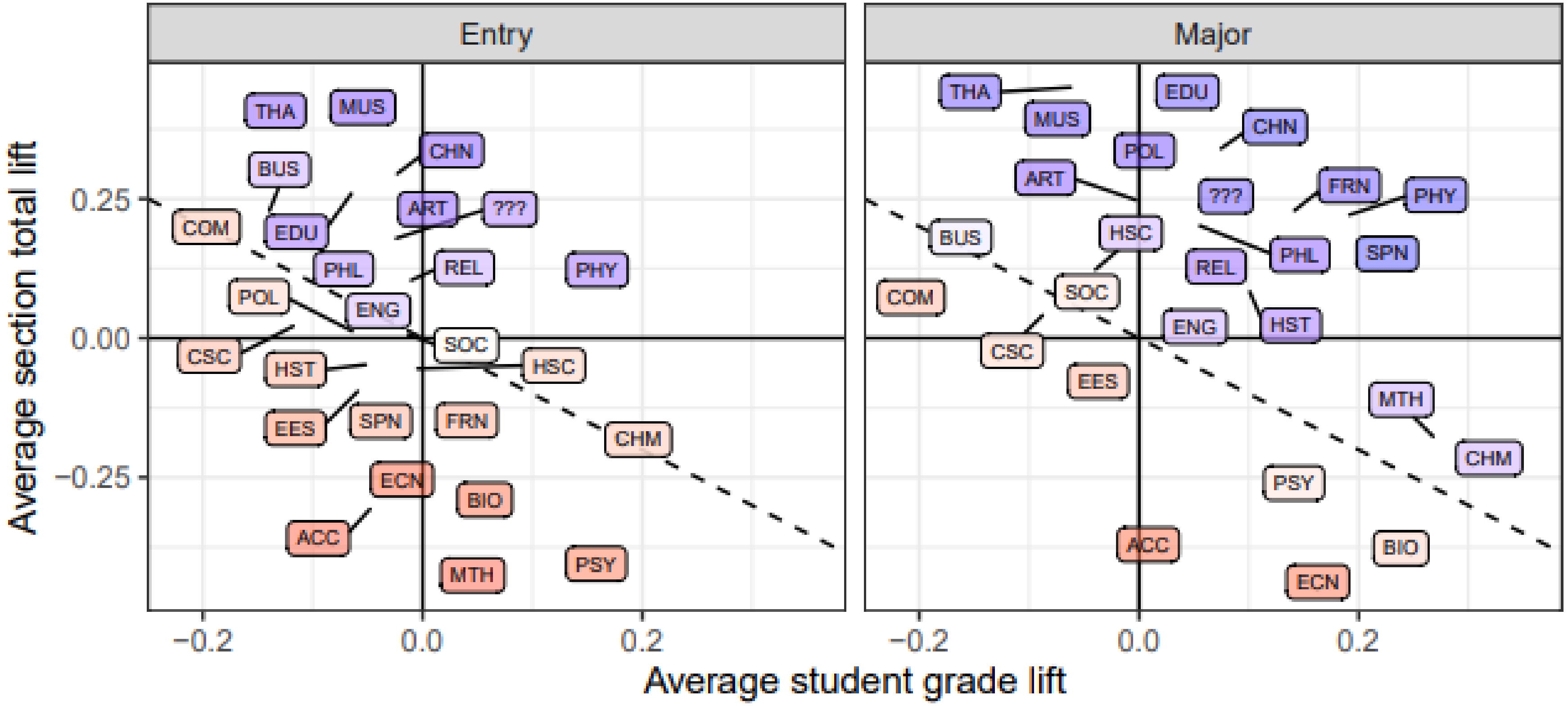
Figure 3: (A) Correlations between subject residuals, with |R| < .10 suppressed and (B) correlations between words in subject catalog descriptions.

Models of Learning Development

Name	N	R2	Intercept	SectionLift	Time	Subject	StudentLift	Interaction
BIO Application	1374	0.43	0.26	0.27	0.15	0.33	0.13	0.38
BIO Graphical Literacy	991	0.38	0.25	0.25	0.23	0.39		0.42
BIO Structure & Function	1098	0.44	0.26	0.25	0.19	0.30	0.08	0.46
COM COM Mediated Messages	869	0.29	0.23	0.15	0.41		0.16	0.19
ECN ECN Analytical Reasoning	1050	0.52		-0.19	0.18	0.51	0.13	0.32
ECN ECN Empirical Application	1009	0.40		-0.16	0.21	0.43	0.16	0.19
ECN ECN Quantitative Methods	639	0.29	0.10		0.17	0.23	0.22	
HST Historical Evidence	839	0.45	0.28	0.20	0.20	0.35	0.27	
HST Historical Methods	868	0.49	0.27	0.37	0.30	0.32	0.37	
MLL Foreign Language Listening Comprehension	1403	0.19	0.33	-0.18	0.13	0.49	0.27	
MLL Foreign Language Oral Proficiency	1522	0.20	0.29	-0.13		0.40	0.31	
MLL Foreign Language Reading Proficiency	1348	0.18	0.32	-0.15		0.47	0.23	
MLL Foreign Language Writing Proficiency	1506	0.23	0.29			0.38	0.36	
MUS Musical Performance	248	0.16	0.36					
MUS Musical Technique	231	0.20	0.30					
U Collaboration	361	0.38	0.10	-0.20	0.34	0.36	0.25	
U Creative/Inductive Thinking	6372	0.38	0.20	0.20	0.32	0.15	0.24	0.09
U Data Analysis	696	0.29	0.17	0.38	0.25	0.23	0.29	
U Discipline Writing	12722	0.39	0.15	0.22	0.42	0.12	0.22	0.11
U Oral Communication	2561	0.34	0.17	0.26	0.32	0.36	0.15	0.12
U Research	935	0.45	0.08	0.11	0.42	0.36	0.20	
U Rules-Based Thinking	6393	0.39	0.18	0.18	0.36	0.19	0.26	
U Scientific Literature	1011	0.40	0.11	0.25	0.50	0.19	0.12	0.23

Table 5: Regression statistics for learning outcomes ratings

Match between Students and Subjects



USNA Study: Course Difficulty and Learning

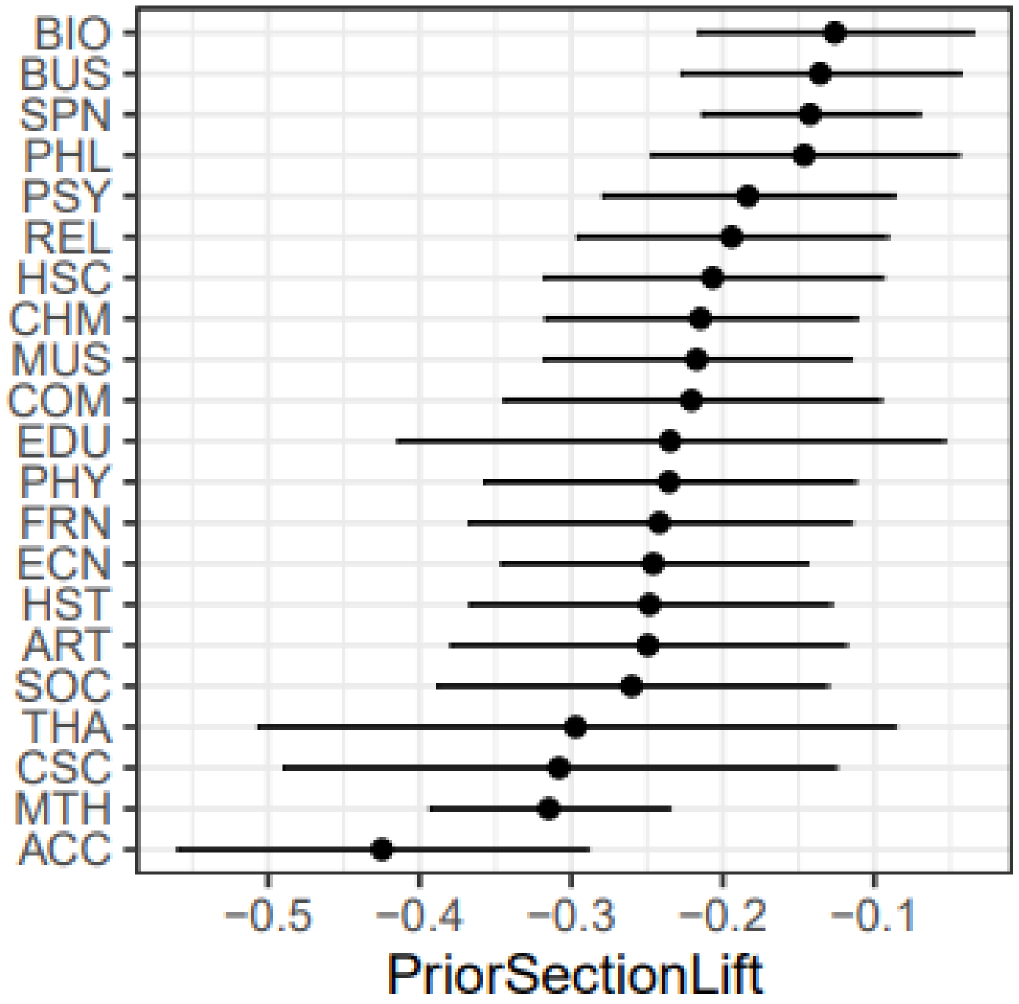
Exploring a variety of mechanisms, we suggest that instructors harm students […] by producing misleading signals regarding the difficulty of the subject and the "soft skills" needed for college success.

Insler, M., McQuoid, A. F., Rahman, A., Smith, K. A. (2021). Fear and Loathing in the Classroom: Why Does Teacher Quality Matter?. *IZA Institute of Labor Economics.* DP 14036.

Course Difficulty and Learning

Table 3: Modeling next grades in the same subject

	Dependent variable:	
	NextGrade	
PriorGrade	0.191***	(0.183, 0.200)
StudentLift	0.839***	(0.828, 0.851)
PriorSectionLift	−0.159***	(−0.176, −0.143)
SubjectLift	0.847***	(0.826, 0.867)
NextSectionLift	1.206***	(1.190, 1.223)
Constant	2.592***	(2.564, 2.619)
Observations	53,051	
R^2	0.645	
Adjusted R^2	0.645	
Residual Std. Error	0.440 (df = 53045)	
F Statistic	19,306.760*** (df = 5; 53045)	
Note:	*p<0.1; **p<0.05; ***p<0.01	



Rethinking Grades

- Grades are averages of direct observations, not “indirect” by any reasonable definition
- Understanding student learning requires a lot of data. It’s unwise to discard the large corpus of grade data without looking at it.
- For a graduate GPA is a reliable measure of subject learning in the major
- Understanding how students develop is a hierarchical problem: working from general to specific is a good approach. Because of their ubiquity, grades can do that in ways that disjoint assessment projects cannot.
- There are many use cases for grade data in combination with all sorts of data. An especially important application is to equity issues.