Relationship between course grades in introductory physics and pre-instruction assessment scores

David E. Meltzer¹ and Dakota H. King²

¹Arizona State University ²University of Arizona College of Veterinary Medicine

Supported in part by NSF DUE #1504986 and #1914712

Acknowledgments

• Diagnostic data have been provided by (among others):

Vince Coletta (Loyola Marymount University)

Steven Pollock (University of Colorado, Boulder)

Christopher Varney (University of West Florida)

Assessment Pretests

- Diagnostic pretest covering pre-college mathematics ("Math")
 - calculators allowed
- Pre-instruction tests of scientific reasoning skill and physics concept knowledge:
 - Lawson Test of Scientific Reasoning ("Lawson")
 - Force Concept Inventory (FCI)
- Why do this? Perhaps ultimately we can offer special assistance to those students who need it most.

Sample Description

- 30 introductory physics classes from 4 universities, over 2000 total students
- Instruction in most classes was "non-traditional," generally highly interactive using research-based instructional materials and methods

Course and Institution Code

Alg-1: Algebra-based course, first semester Alg-2: Algebra-based course, second semester Calc-1: Calculus-based course, first semester Calc-2: Calculus-based course, second semester

ASU-P: Arizona State University, Polytechnic campus

ASU-T: Arizona State University, Tempe campus

LMU: Loyola Marymount University

UWF: University of West Florida

CU: University of Colorado, Boulder

Comparing probabilities of high* and low* grades

- What is the probability of a student with a *high* score on a preinstruction assessment getting a high grade in the class?
- How does that compare to a *low-scoring* student's probability of getting a high grade?

(and, same questions for probabilities of getting a low grade)

*In this context, "high" and "low" mean "top quartile" and "bottom quartile"

Consistent result:

High scorers on the diagnostic pretests were much more likely to get *high* grades than were low scorers

High Course Grade vs. Mathematics Diagnostic Pretest Score

Course	Campus	Ν	Top-quartile Math: % with top-quartile grades	Bottom-quartile Math: % with top-quartile grades	High-grade odds ratio
Alg-1 2021a	ASU-P	39	51%	10%	5.0
Alg-1 2021b	ASU-P	42	44%	10%	4.6
Alg-1 2022a	ASU-P	40	27%	6%	4.4
Alg-1 2022b	ASU-P	52	49%	10%	5.1
Alg-1 2023a	ASU-P	42	39%	10%	4.1
Alg-1 2023b	ASU-P	46	64%	9%	7.3
Alg-2 2022	ASU-P	75	46%	21%	2.2
Alg-2 2023	ASU-P	92	41%	13%	3.2
Alg-2 2024	ASU-P	99	51%	8%	6.1
Alg-2 2021	ASU-T	129	30%	39%	0.8
Calc-1 2021a	UWF	53	43%	0%	"∞"
Calc-1 2021b	UWF	42	43%	0%	"∞"
Calc-2 2021	UWF	58	43%	14%	3.1
AVERAGE	(unweighted)	(809)	44%	12%	3.8

High Course Grade vs. Mathemati	cs Diagnostic Pretest Score

Course	Campus	N	Top-quartile Math: % with top-quartile grades	Bottom-quartile Math: % with top-quartile grades	High-grade odds ratio
Alg-1 2021a	ASU-P	39	51%	10%	5.0
Alg-1 2021b	ASU-P	42	44%	10%	4.6
Alg-1 2022a	ASU-P	4 0	27%	6%	4.4
Alg-1 2022b	ASU-P	52	49%	10%	5.1
Alg-1 2023a	ASU-P	42	39%	10%	4.1
Alg-1 2023b	ASU-P	46	64%	9%	7.3
Alg-2 2022	ASU-P	75	46%	21%	2.2
Alg-2 2023	ASU-P	92	41%	13%	3.2
Alg-2 2024	ASU-P	99	51%	8%	6.1
Alg-2 2021	ASU-T	129	30%	39%	0.8
Calc-1 2021a	UWF	53	43%	0%	"∞"
Calc-1 2021b	UWF	42	43%	0%	"∞"
Calc-2 2021	UWF	58	43%	14%	3.1
AVERAGE	(unweighted)	(809)	44%	12%	3.8

High Course Grade vs. Mathematics Diagnostic Pretest Score

Course	Campus	N	Top-quartile Math: % with top-quartile grades	Bottom-quartile Math: % with top-quartile grades	High-grade odds ratio
Alg-1 2021a	ASU-P	39	51%	10%	5.0
Alg-1 2021b	ASU-P	42	44%	10%	4.6
Alg-1 2022a	ASU-P	40	27%	6%	4.4
Alg-1 2022b	ASU-P	52	49%	10%	5.1
Alg-1 2023a	ASU-P	42	39%	10%	4.1
Alg-1 2023b	ASU-P	46	64%	9%	7.3
Alg-2 2022	ASU-P	75	46%	21%	2.2
Alg-2 2023	ASU-P	92	41%	13%	3.2
Alg-2 2024	ASU-P	99	51%	8%	6.1
Alg-2 2021	ASU-T	129	30%	39%	0.8
Calc-1 2021a	UWF	53	43%	0%	"∞"
Calc-1 2021b	UWF	42	43%	0%	"∞"
Calc-2 2021	UWF	58	43%	14%	3.1
AVERAGE	(unweighted)	(809)	44%	12%	3.8

	High Course Grade vs. Mathematics Diagnostic Pretest Score							
Course	Campus	Ν	Top-quartile Math: % with top-quartile grades	Bottom-quartile Math: % wi top-quartile grades	th High-grade odds ratio			
Alg-1 2021a	ASU-P	39	51%	10%	5.0			
Alg-1 2021b	ASU-P	42	44%	10%	4.6			
Alg-1 2022a	ASU-P	40	27%	6%	4.4			
Alg-1 2022b	ASU-P	52	49%	10%	5.1			
Alg-1 2023a	ASU-P	42	39%	10%	4.1			
Alg-1 2023b	ASU-P	46	64%	9%	7.3			
Alg-2 2022	ASU-P	75	46%	21%	2.2			
Alg-2 2023	ASU-P	92	41%	13%	3.2			
Alg-2 2024	ASU-P	99	51%	8%	6.1			
Alg-2 2021	ASU-T	129	30%	39%	0.8			
Calc-1 2021a	UWF	53	43%	0%	"∞"			
Calc-1 2021b	UWF	42	43%	0%	"∞"			
Calc-2 2021	UWF	58	43%	14%	3.1			
AVERAGE	(unweighted)	(809)	44%	12%	3.8			

High Course Grade vs. Mathematics Diagnostic Pretest Score

Alg-1 2021b AS	SU-P 4	39 12	51%	10%	5.0
		12			
Alg-1 2022a AS	SU-P 4		44%	10%	4.6
		40	27%	6%	4.4
Alg-1 2022b AS	SU-P 5	52	49%	10%	5.1
Alg-1 2023a AS	SU-P 4	12	39%	10%	4.1
Alg-1 2023b AS	SU-P 4	46	64%	9%	7.3
Alg-2 2022 AS	SU-P 7	75	46%	21%	2.2
Alg-2 2023 AS	SU-P 9	92	41%	13%	3.2
Alg-2 2024 AS	SU-P 9	99	51%	8%	6.1
Alg-2 2021 AS	SU-T 12	29	30%	39%	0.8
Calc-1 2021a UV	WF 5	53	43%	0%	"∞"
Calc-1 2021b UV	WF 4	12	43%	0%	"∞"
Calc-2 2021 UV	WF 5	58	43%	14%	3.1
AVERAGE (ur	nweighted) (8(09)	44%	12%	3.8

High Course Grade vs. Mathematics Diagnostic Pretest Score							
Campus	N	Top-quartile Math: % with top-quartile grades	Bottom-quartile Math: % with top-quartile grades	High-grade odds ratio			



High scorers on math pretest were 3.8 times more likely to get a high grade than were low scorers

High Course Grade vs. Lawson Test of Scientific Reasoning Pretest Score

Course	Campus	N	Top-quartile Lawson: % with top-quartile grades	Bottom-quartile Lawson: % with top- quartile grades	High-grade odds ratio
Alg-1 2021a	ASU-P	35	46%	23%	2.0
Alg-1 2021b	ASU-P	38	32%	8%	4.0
Alg-1 2022a	ASU-P	41	49%	10%	5.0
Alg-1 2022b	ASU-P	54	57%	10%	5.6
Alg-1 2023a	ASU-P	36	39%	33%	1.2
Alg-1 2023b	ASU-P	44	55%	9%	6.0
Alg-2 2022	ASU-P	73	41%	6%	7.6
Alg-2 2023	ASU-P	92	52%	10%	5.0
Alg-2 2024	ASU-P	90	42%	5%	9.2
Alg-1	CU	469	45%	8%	5.5
Calc-2	CU	276	57%	8%	6.9
Alg-1 2007	LMU	24	50%	0%	"∞"
Alg-1 2009	LMU	51	34%	11%	3.2
Alg-1 2011	LMU	57	53%	18%	2.9
Alg-1 2012	LMU	44	64%	6%	10.5
Alg-1 2013	LMU	30	53%	12%	4.6
Alg-1 2014	LMU	33	61%	0%	"∞"
Alg-1 2015	LMU	24	63%	0%	"∞"
Alg-1 2016	LMU	35	41%	0%	"∞"
Alg-1 2018	LMU	47	54%	9%	6.3
Alg-1 2021	LMU	27	44%	0%	"∞"
AVERAGE	(unweighted)	(1620)	49%	9%	5.5

High Course Grade vs. FCI						
Course	Campus	N	Top-quartile FCI: % with top-quartile grades	Bottom-quartile FCI: % with top-quartile grades	High-grade odds ratio	
Alg-1 2018	ASU-P	48	40%	8%	4.8	
Alg-1 2019	ASU-P	63	38%	13%	3.0	
Alg-1 2021a	ASU-P	35	57%	0%	"∞"	
Alg-1 2021b	ASU-P	37	32%	17%	1.9	
Alg-1 2022a	ASU-P	41	21%	15%	1.4	
Alg-1 2022b	ASU-P	52	26%	7%	3.9	
Alg-1 2023a	ASU-P	40	30%	20%	1.3	
Alg-1 2023b	ASU-P	47	55%	18%	3.1	
Alg-1	CU	470	41%	12%	3.5	
Alg-1 2007	LMU	23	87%	0%	"∞"	
Alg-1 2009	LMU	51	63%	0%	"∞"	
Alg-1 2012	LMU	44	50%	0%	"∞"	
Alg-1 2013	LMU	30	51%	0%	"∞"	
Alg-1 2014	LMU	33	43%	12%	3.6	
Alg-1 2015	LMU	24	67%	0%	"∞"	
Alg-1 2016	LMU	34	71%	0%	"∞"	
Alg-1 2018	LMU	47	34%	14%	2.4	
Alg-1 2021	LMU	27	44%	0%	"∞"	
Calc-1 2012	ASU-P	40	43%	0%	" _∞ "	
Calc-1 2013a	ASU-P	18	44%	0%	" _∞ "	
Calc-1 2013b	ASU-P	48	54%	17%	3.3	
Calc-1 2021a	UWF	62	29%	26%	1.1	
Calc-1 2021b	UWF	53	40%	15%	2.6	
AVERAGE	(unweighted)	(1367)	46%	8%	5.4	

<

- High scorers on the pretests were about 5 times more likely to get a high grade than low scorers.
- Low scorers on the pretests were about *4 times* more likely to get a low grade than high scorers.

High and low grades for high and low scorers were compared for more than 2000 students in 30 distinct classes at 4 universities, yielding a total of 114 high/low comparisons.

> The quartile ratios were greater than 1.0 in 111 of the 114 cases (97%).

Relevant Questions

- Which, if any, of the diagnostic pretests is most predictive of students' performance?
- Does using multiple predictor variables offer greater predictive power than using just one of them?
- Can an "accurate" predictive model be created that incorporares multiple predictor variables?
- Does better performance on one pretest indicate that another pretest is more (or less) predictive? (This would be an "interaction" effect.)

Relevant Questions

• Which, if any, of the diagnostic pretests is most predictive of students' performance? *Varies with the course*

- Courses and instructors differ on the relative emphasis placed on conceptual problems, mathematical problem solving, and problems requiring significant reading and reasoning skills.
- There are also many possible ways to compare relative "predictability"
 - ...for example, compare high/low grade ratios for different pretests

High Grade Odds Ratios

	Course	Campus	N (# classes)	High-grade odds ratio, Math (average)	High-grade odds ratio, Lawson (average)	High-grade odds ratio, FCI (average)
\langle	Alg-1	ASU-P	6	5.0	3.0	3.6
	Alg-1	CU	1		5.5	3.5
	Alg-1	LMU	9		12.2	19.6
	Alg-2	ASU-P	3	3.3	6.4	
	Calc-1	UWF	2	"∞"		1.7

High Grade Odds Ratios

Course	Campus	N (# classes)	High-grade odds ratio, Math (average)	High-grade odds ratio, Lawson (average)	High-grade odds ratio, FCI (average)
Alg-1	ASU-P	6	5.0	3.0	3.6
Alg-1	CU	1	\smile	5.5	3.5
Alg-1	LMU	9		12.2	(19.6)
Alg-2	ASU-P	3	3.3	6.4	$_{}$
Calc-1	UWF	2	("∞")		1.7

Low Grade Odds Ratios

Course	Campus	N (# classes)	Low-grade odds ratio, Math (average)	Low-grade odds ratio, Lawson (average)	Low-grade odds ratio, FCI (average)
Alg-1	ASU-P	6	2.8	3.9	3.2
Alg-1	CU	1		(4.4)	1.1
Alg-1	LMU	9		6.8	5.4
Alg-2	ASU-P	3	3.7	2.6	
Calc-1	UWF	2	4.2		3.1

Another Approach: Multiple Regression

• Fit an equation including all predictor variables to the data using ordinary least squares, e.g.:

```
Grades = \beta_0 + \beta_1^* Lawson Pretest + \beta_2^* Math Pretest + \beta_3^* FCI Pretest
```

- Problem: most sample sizes too small to yield significant results, and too much between-class variation to combine samples
- Partial solution:
 - choose one very large sample (Alg-1 CU; N = 466) to compare Lawson and FCI
 - combine three very similar classes taught by same instructor (Alg-2 ASU-P, 2022-23-24; N = 216) to compare Lawson and Math

Reminder: Results for High Grade Odds Ratios for CU sample

Course	Campus	N	High-grade odds ratio, Math	High-grade odds ratio, Lawson	High-grade odds ratio, FCI
Alg-1	CU	466		5.5	3.5

Low Grade Odds Ratios

Course	Campus	N	Low-grade odds ratio, Math	Low-grade odds ratio, Lawson	Low-grade odds ratio, FCI
Alg-1	CU	466		(4.4)	1.1

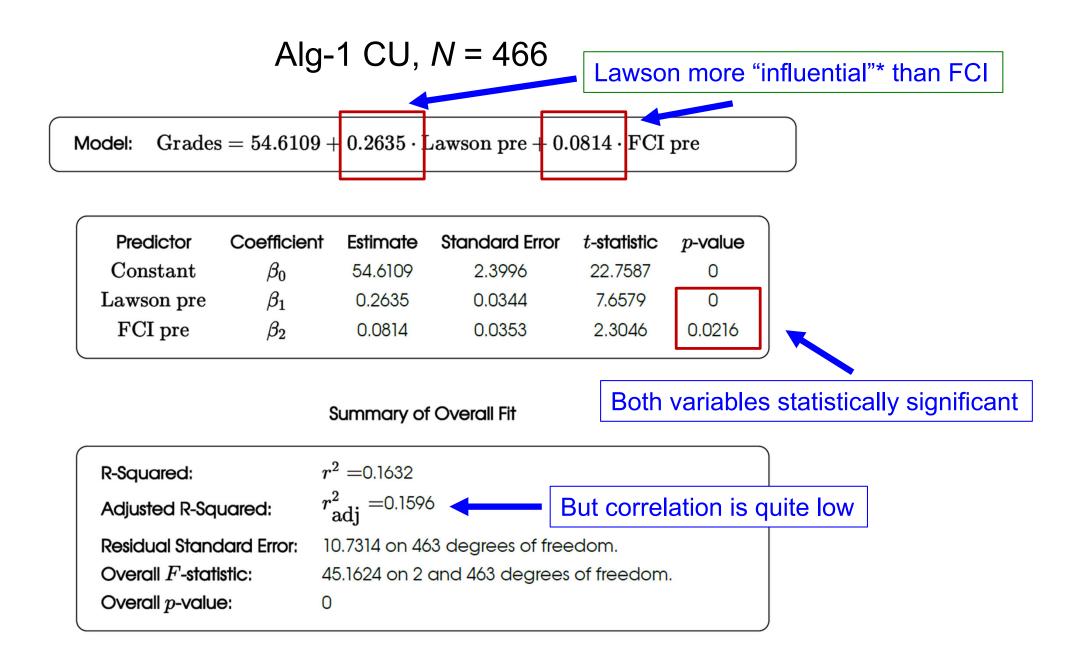
Lawson pretest score seems to be more predictive than FCI pretest score

Alg-1 CU, *N* = 466

Results of Multiple Regression

Alg-1 CU, *N* = 466

Model: Grades = $54.6109 + 0.2635 \cdot \text{Lawson pre} + 0.0814 \cdot \text{FCI pre}$



*weighted more heavily when predicting grades

High Grade Odds Ratios for Alg-2 ASU-P

Course	Campus	N	High-grade odds ratio, Math	High-grade odds ratio, Lawson	High-grade odds ratio, FCI
Alg-2	ASU-P, 2022-23-24	216	2.8	4.1	

High Grade Odds Ratios for Alg-2 ASU-P

Course	Campus	N	High-grade odds ratio, Math	High-grade odds ratio, Lawson	High-grade odds ratio, FCI
Alg-2	ASU-P, 2022-23-24	216	2.8	4.1	

High Grade Odds Ratios for Alg-2 ASU-P

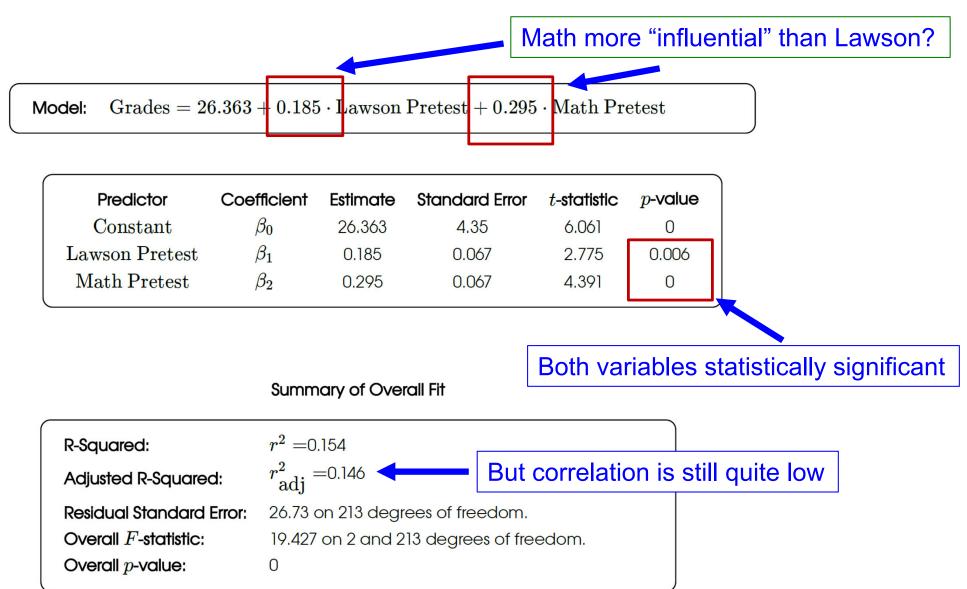
Course	Campus	N	High-grade odds ratio, Math	High-grade odds ratio, Lawson	High-grade odds ratio, FCI
Alg-2	ASU-P, 2022-23-24	216	2.8	4.1	

Low Grade Odds Ratios for Alg-2 ASU-P

Course	Campus	N	Low-grade odds ratio, Math	Low-grade odds ratio, Lawson	Low-grade odds ratio, FCI
Alg-2	ASU-P, 2022-23-24	216	4.8	2.4	

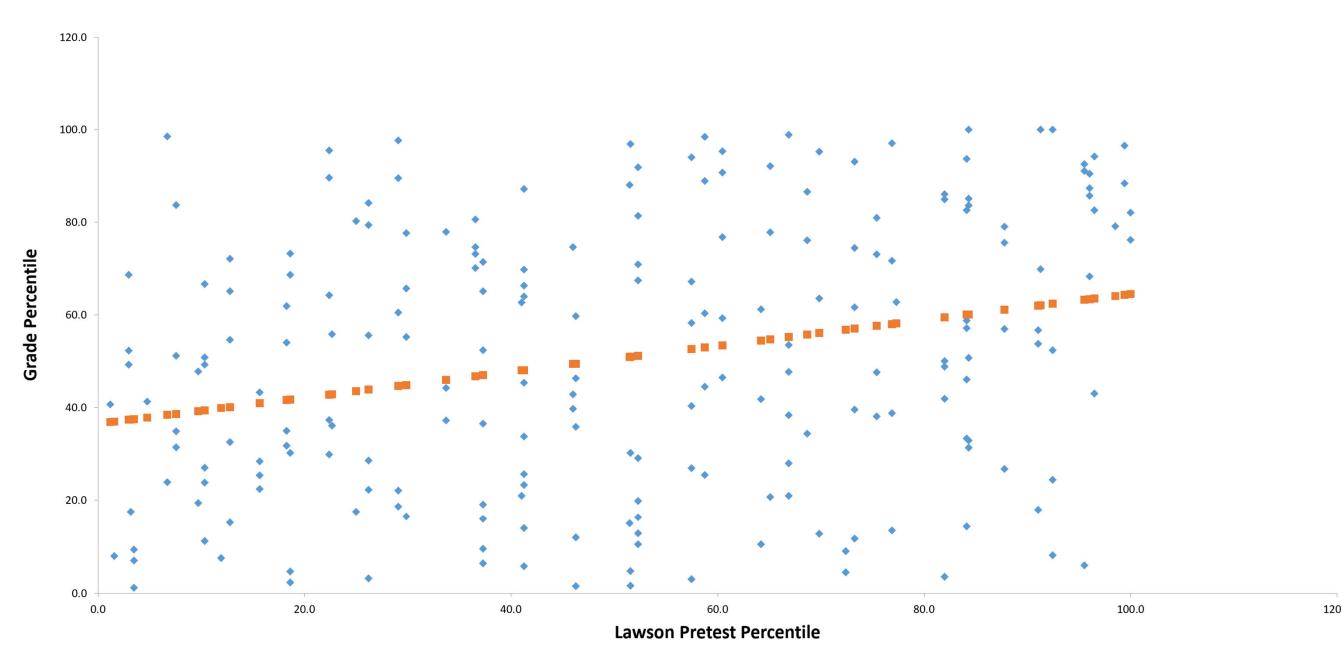


Alg-2 ASU-P, 2022-23-24, *N* = 216

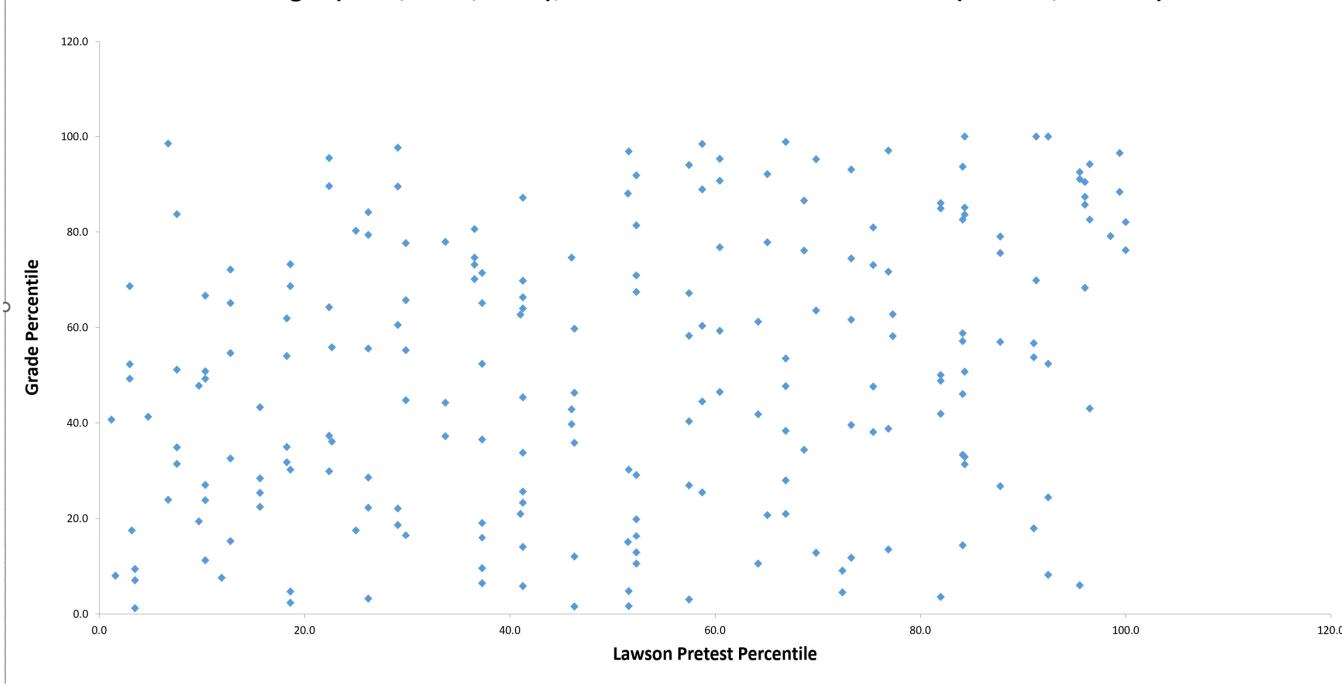


Regression analysis can be misleading

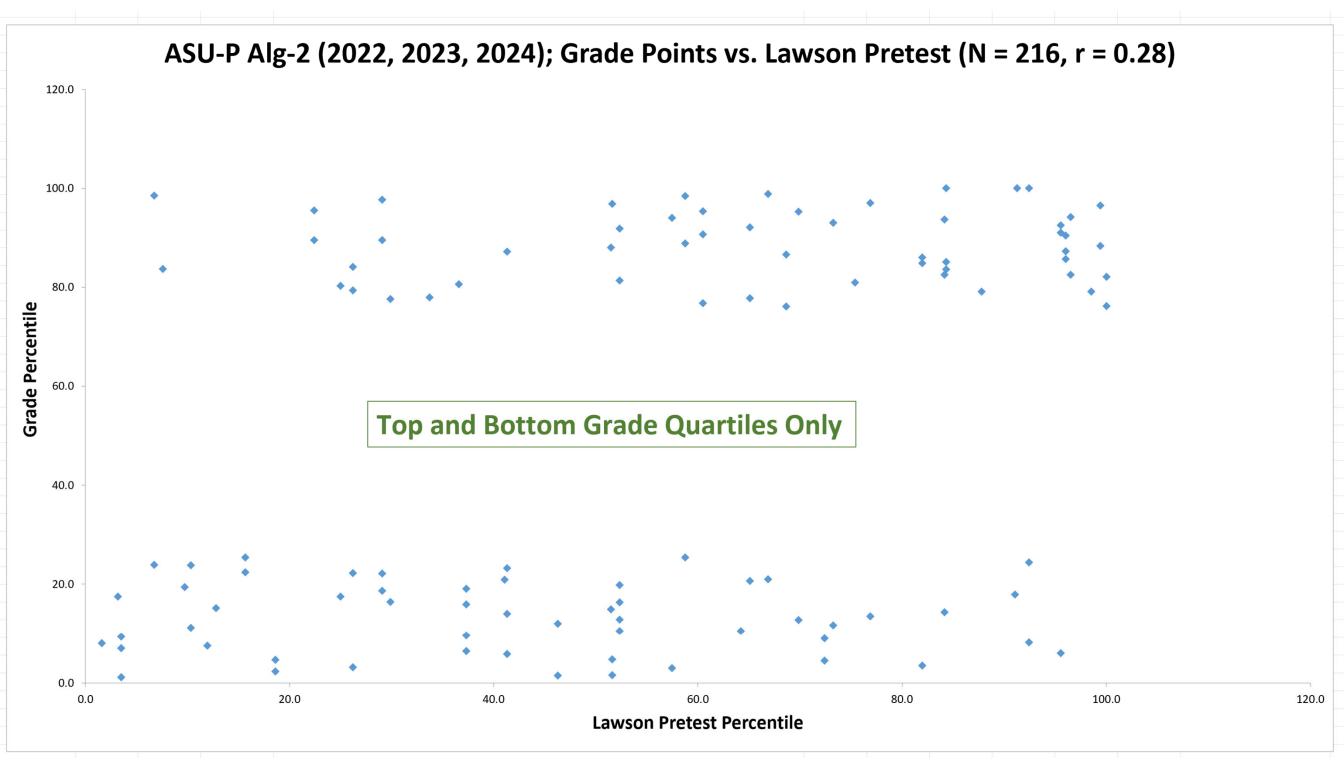
- High scatter in the data leads to relatively low correlation
- However, quartile comparison can reveal highly significant differences between low and high scorers

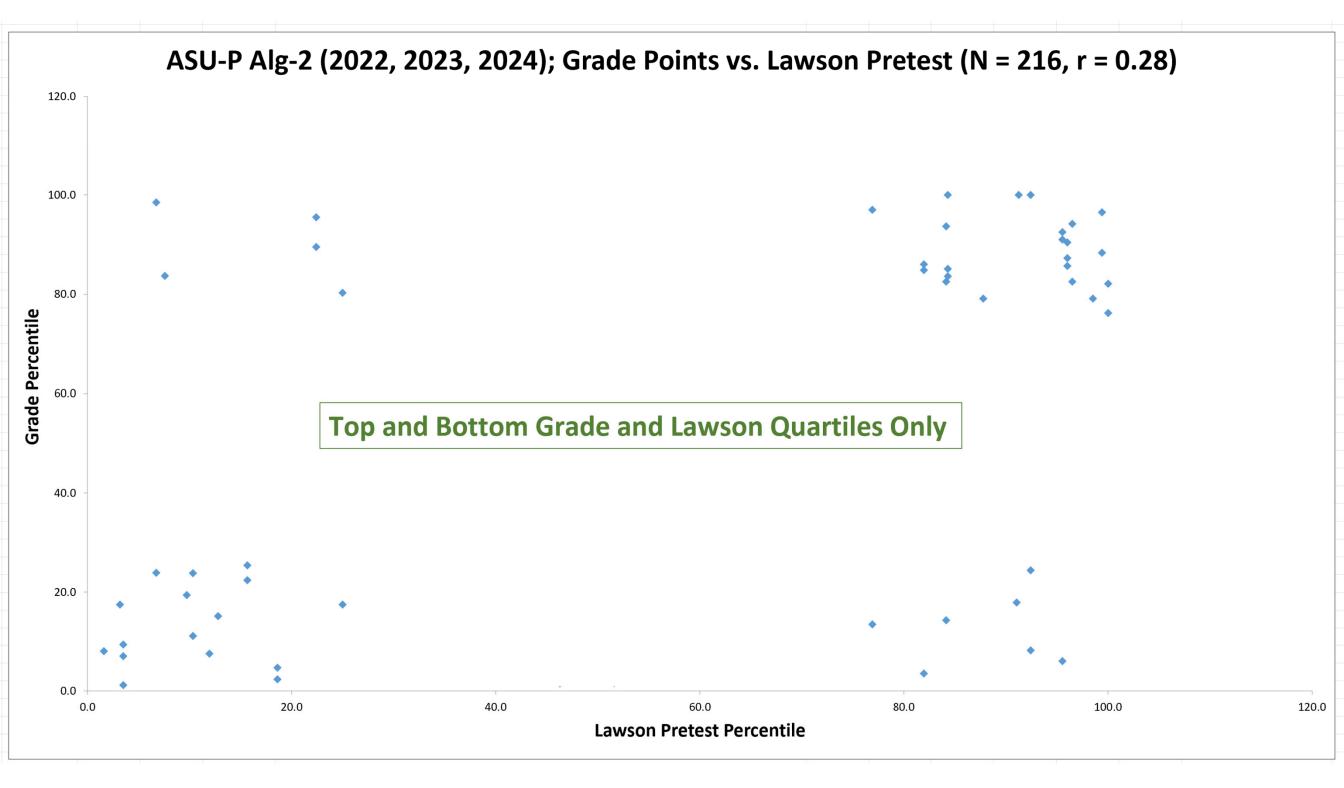


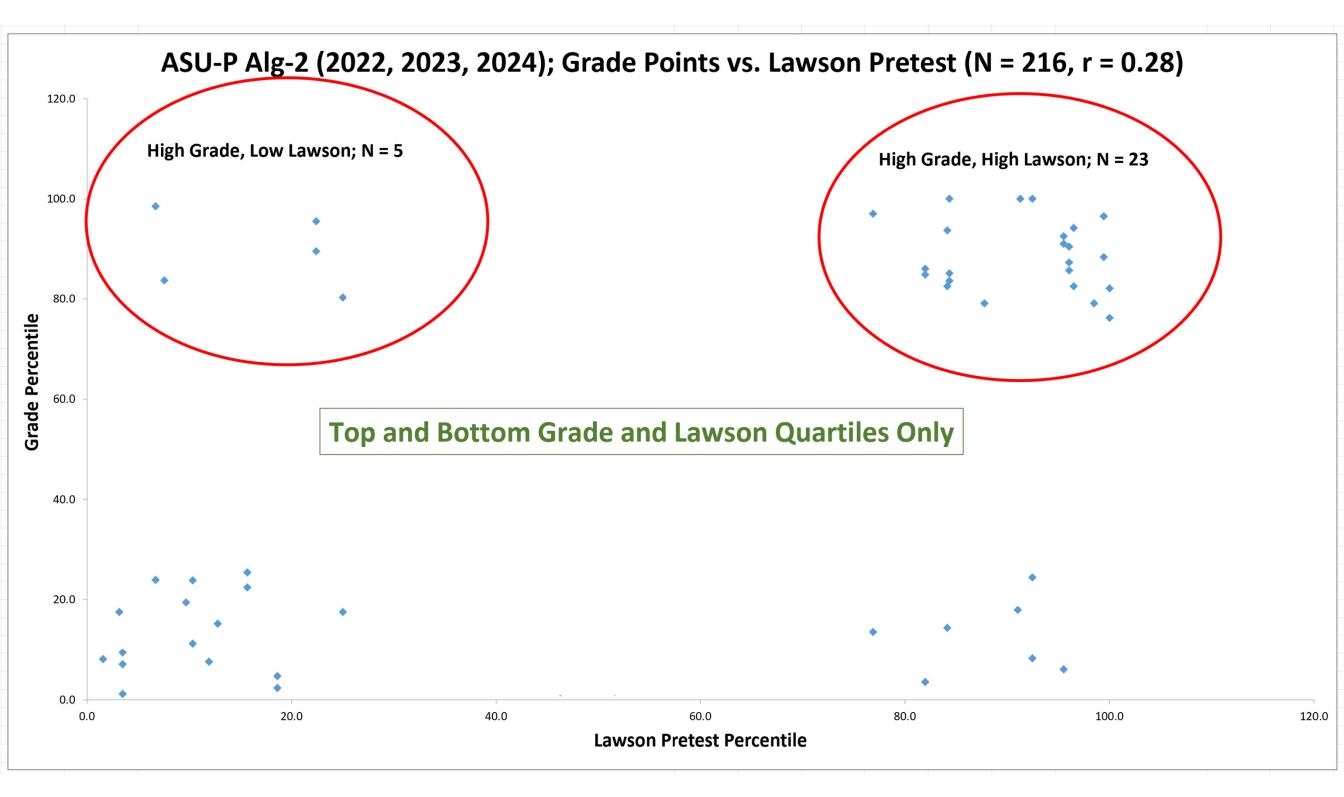
ASU-P Alg-2 (2022, 2023, 2024); Grade Points vs. Lawson Pretest (N = 216, r = 0.28)

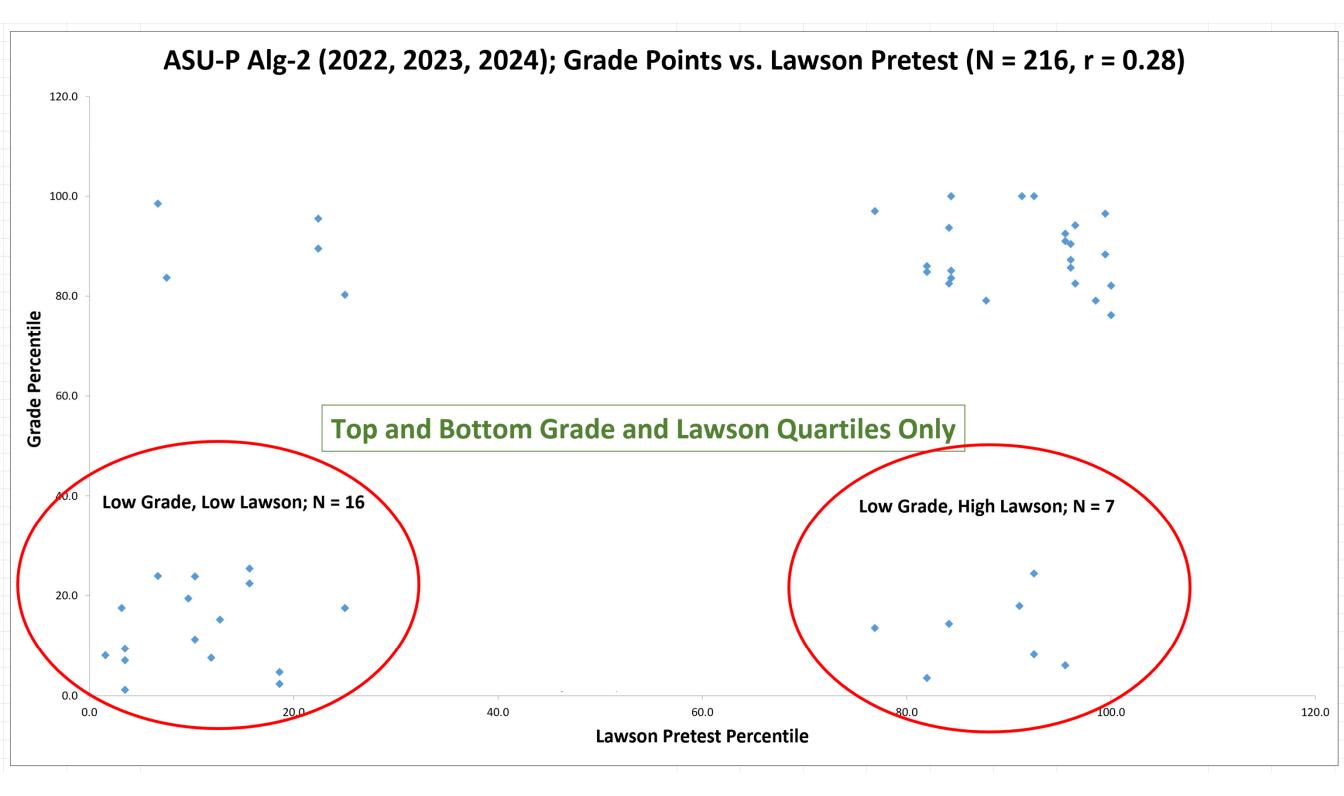


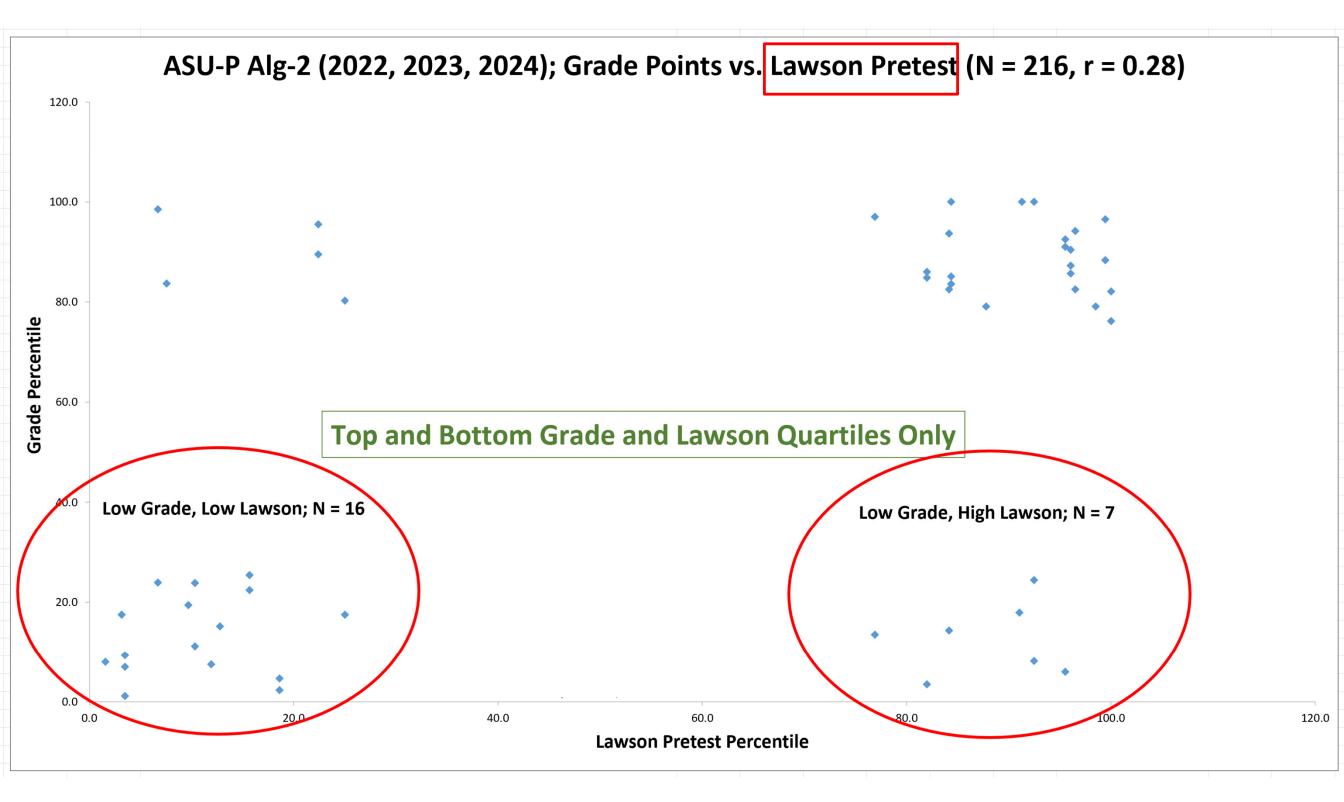
ASU-P Alg-2 (2022, 2023, 2024); Grade Points vs. Lawson Pretest (N = 216, r = 0.28)

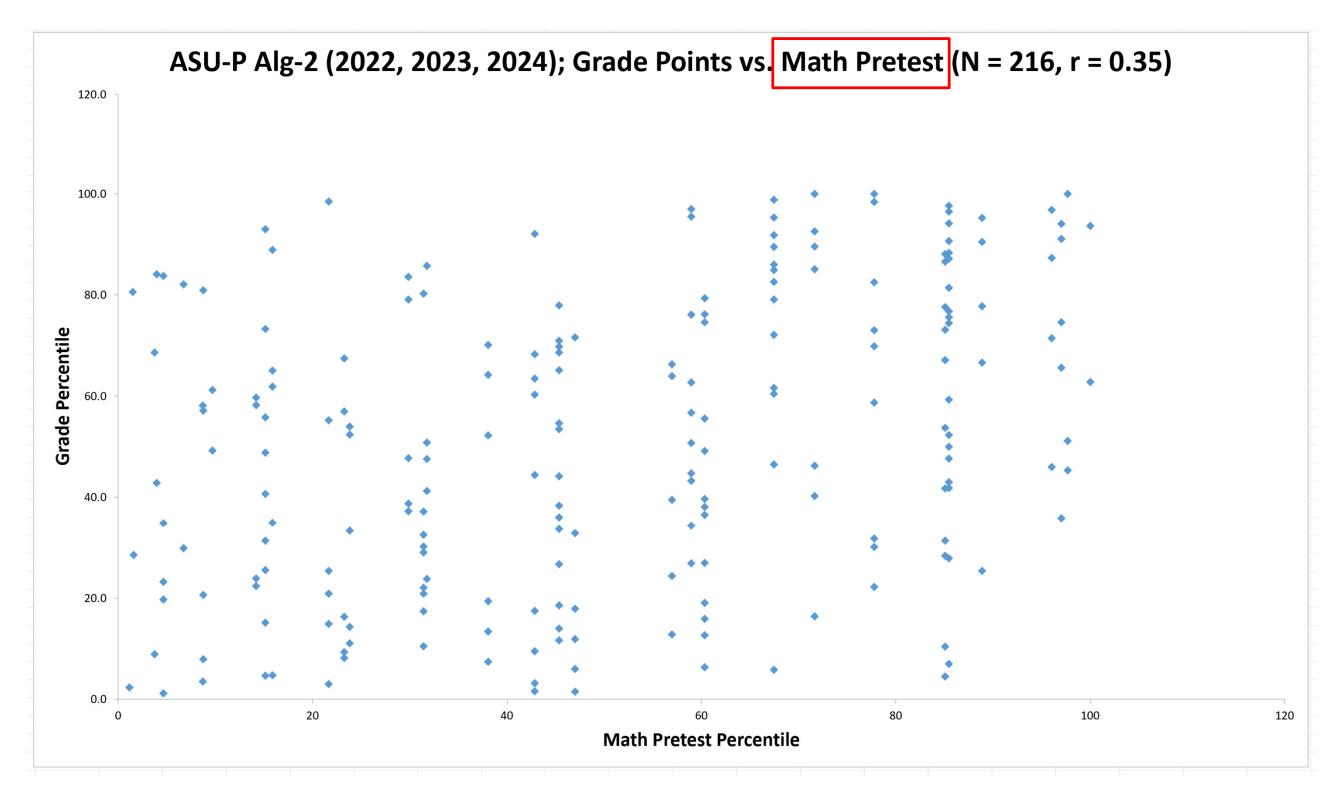


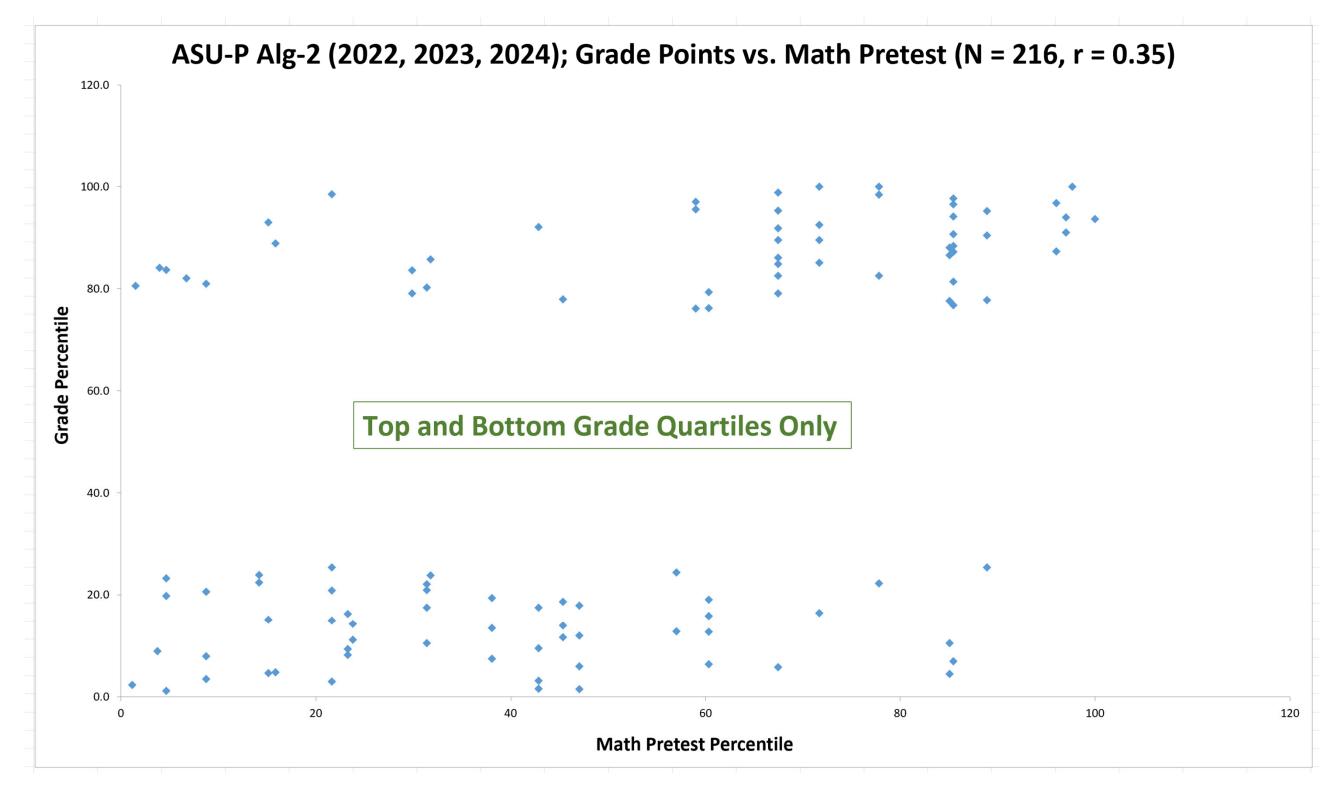


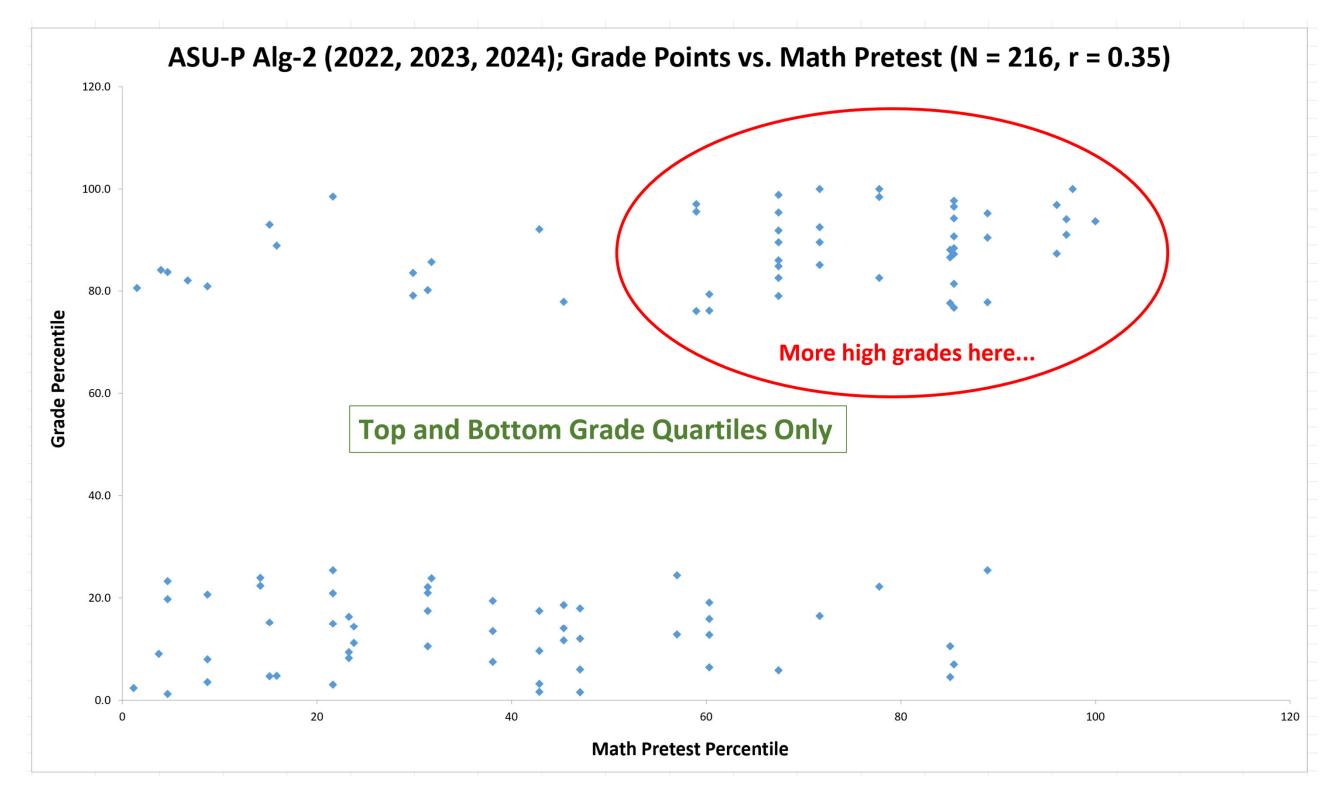


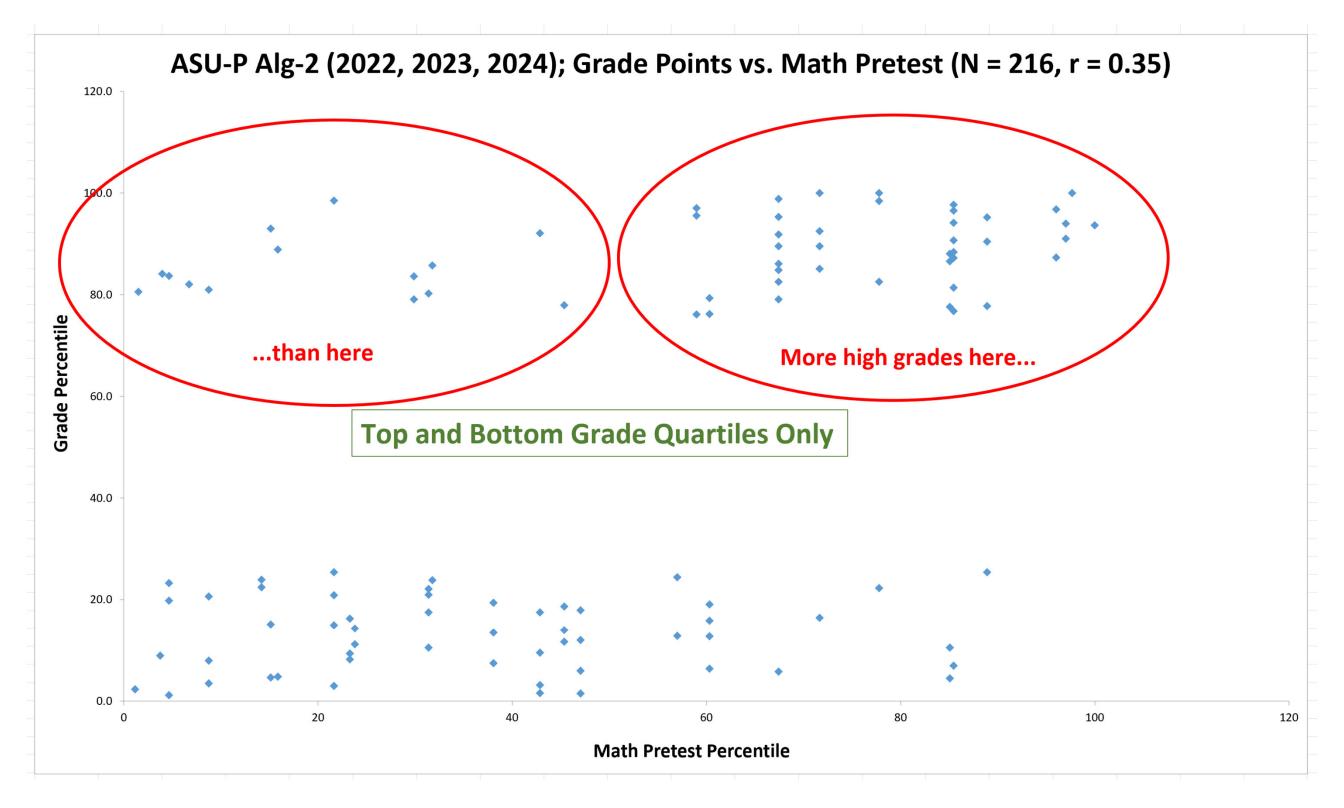


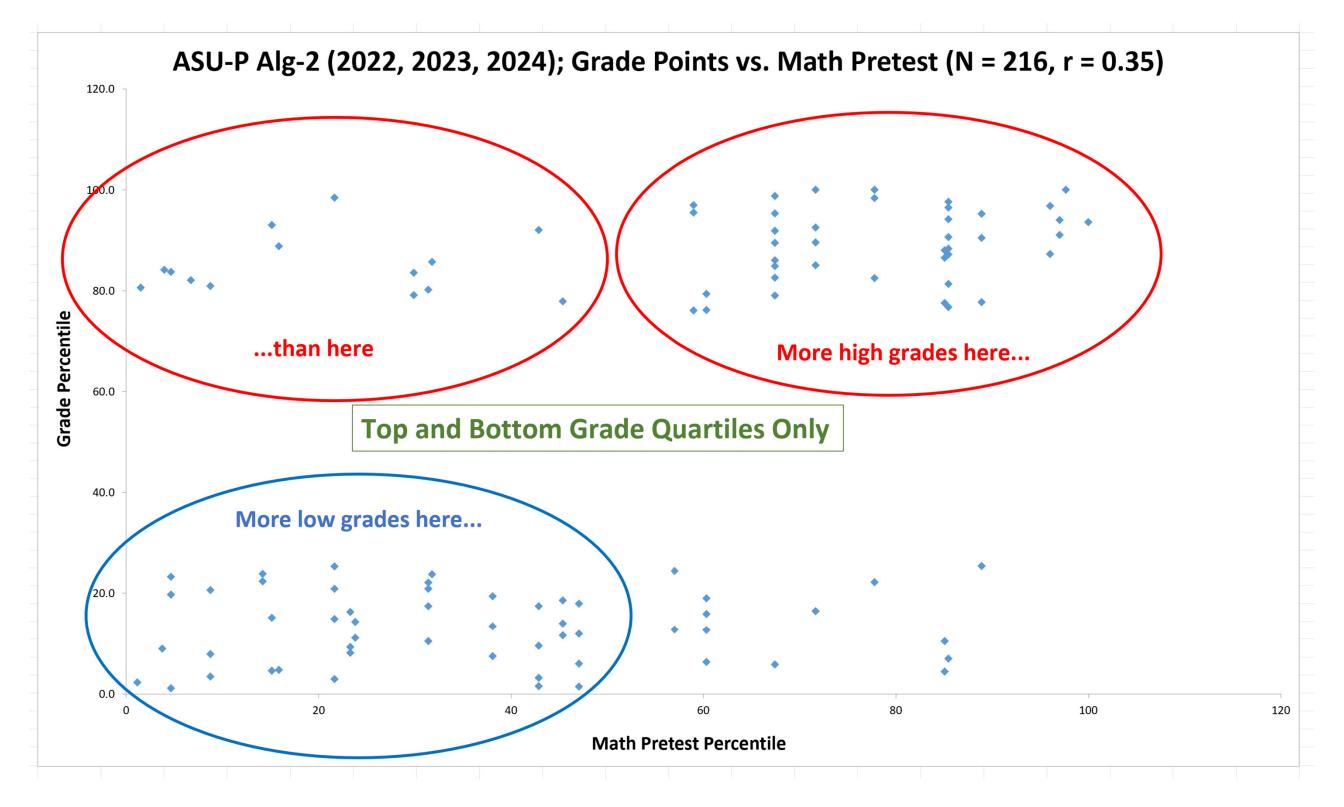


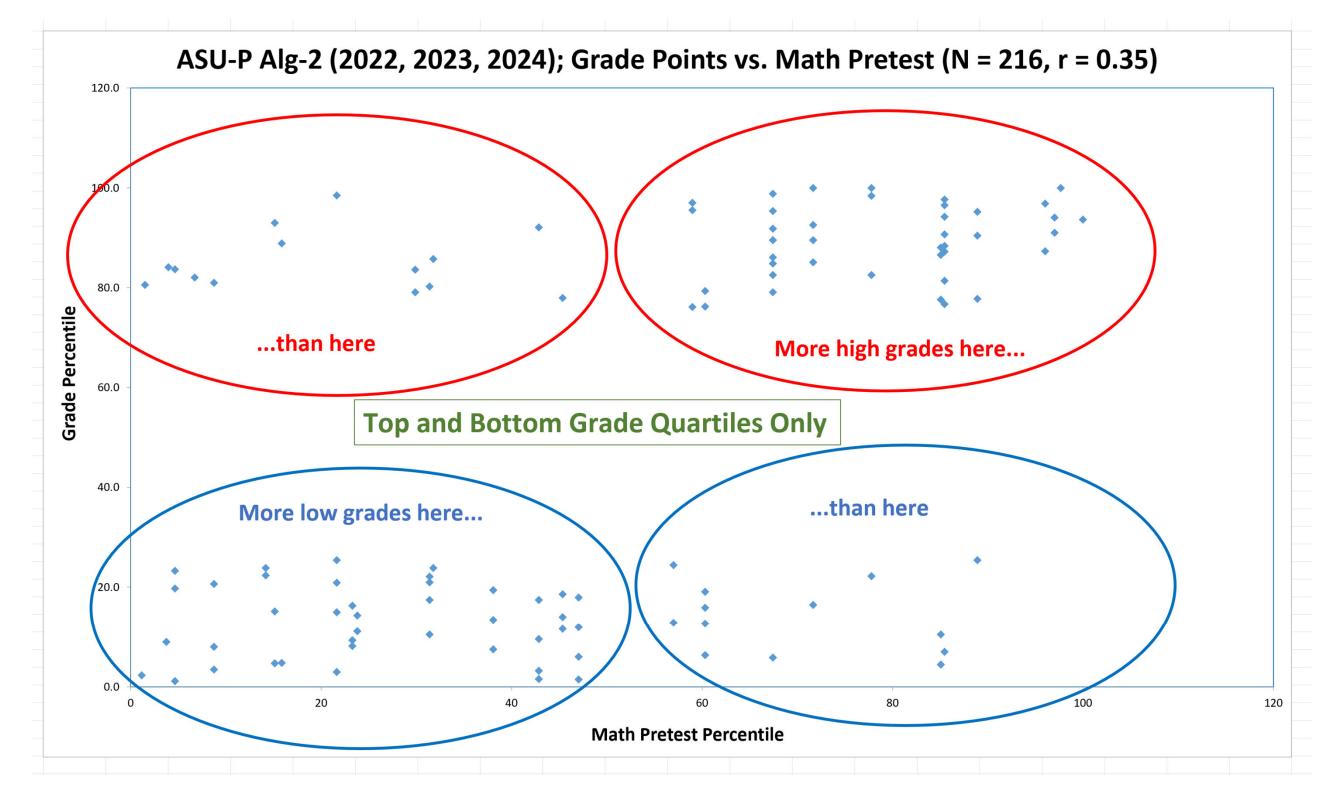












Relevant Questions

Can an "accurate" predictive model be created that incorporares multiple predictor variables? No, but better than random

Alg-2 ASU-P, 2022-23-24, *N* = 216

Model: Grades = $26.363 + 0.185 \cdot \text{Lawson Pretest} + 0.295 \cdot \text{Math Pretest}$

Q: How accurately can this model predict students' grades?

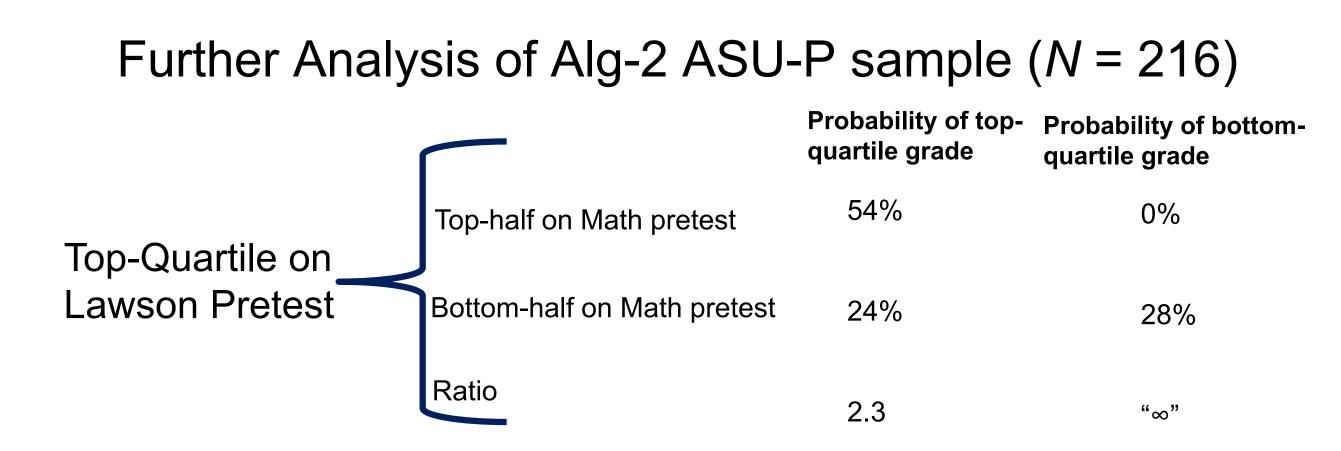
A: If we randomly guess in which grade quartile each student will end up, we'd be right 25% of the time.

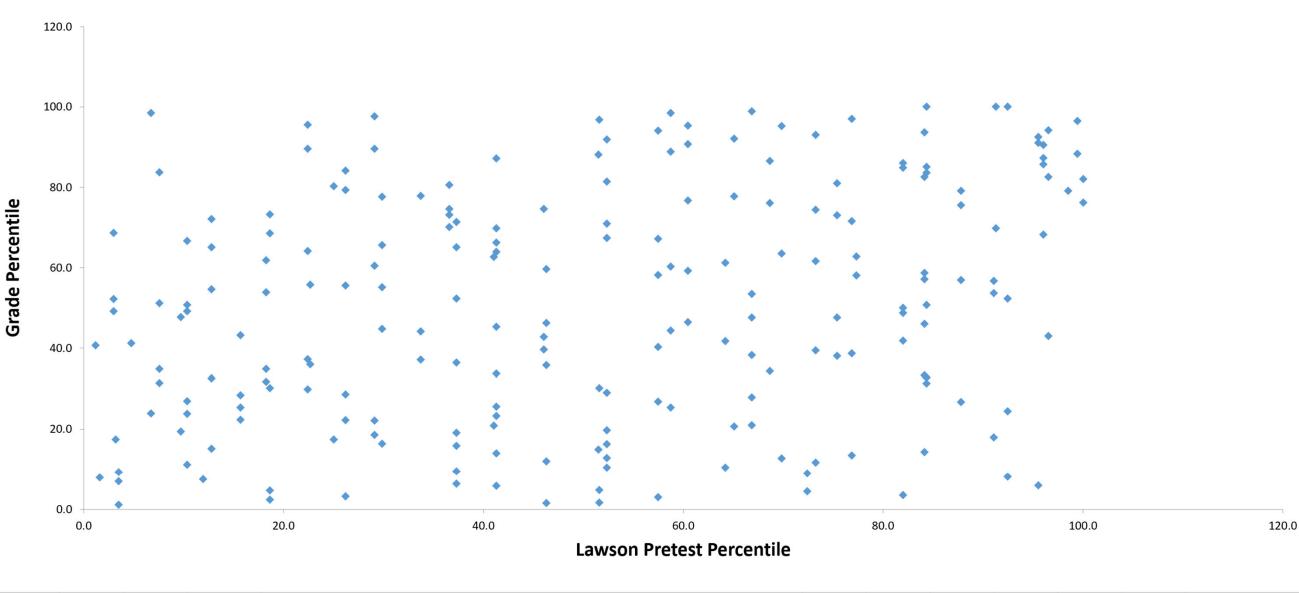
If instead we had applied this model to each *actual** student's Lawson and Math pretest scores, we would have correctly predicted whether they ended up with top- or bottom-quartile grades 45% of the time.

*that is, the students in Alg-2 ASU-P 2022-23-24

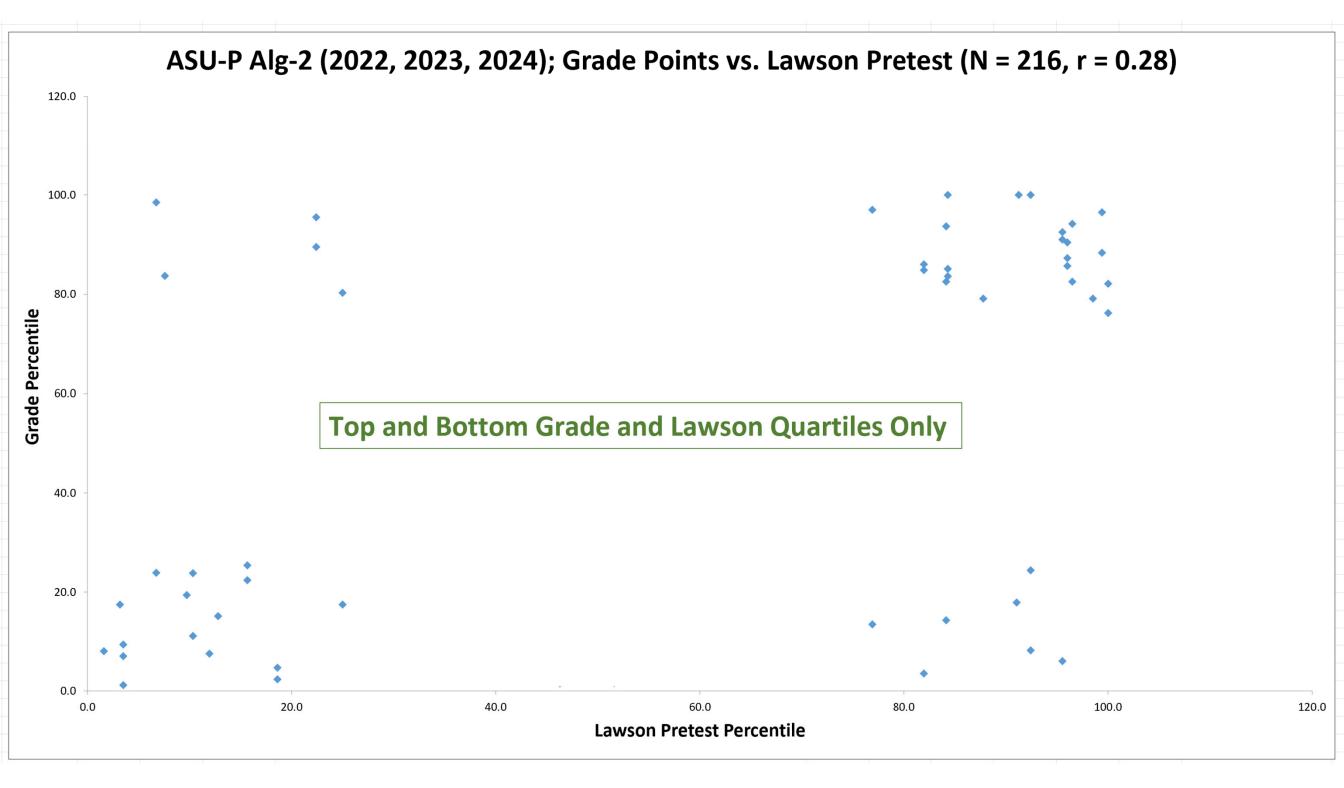
Relevant Questions

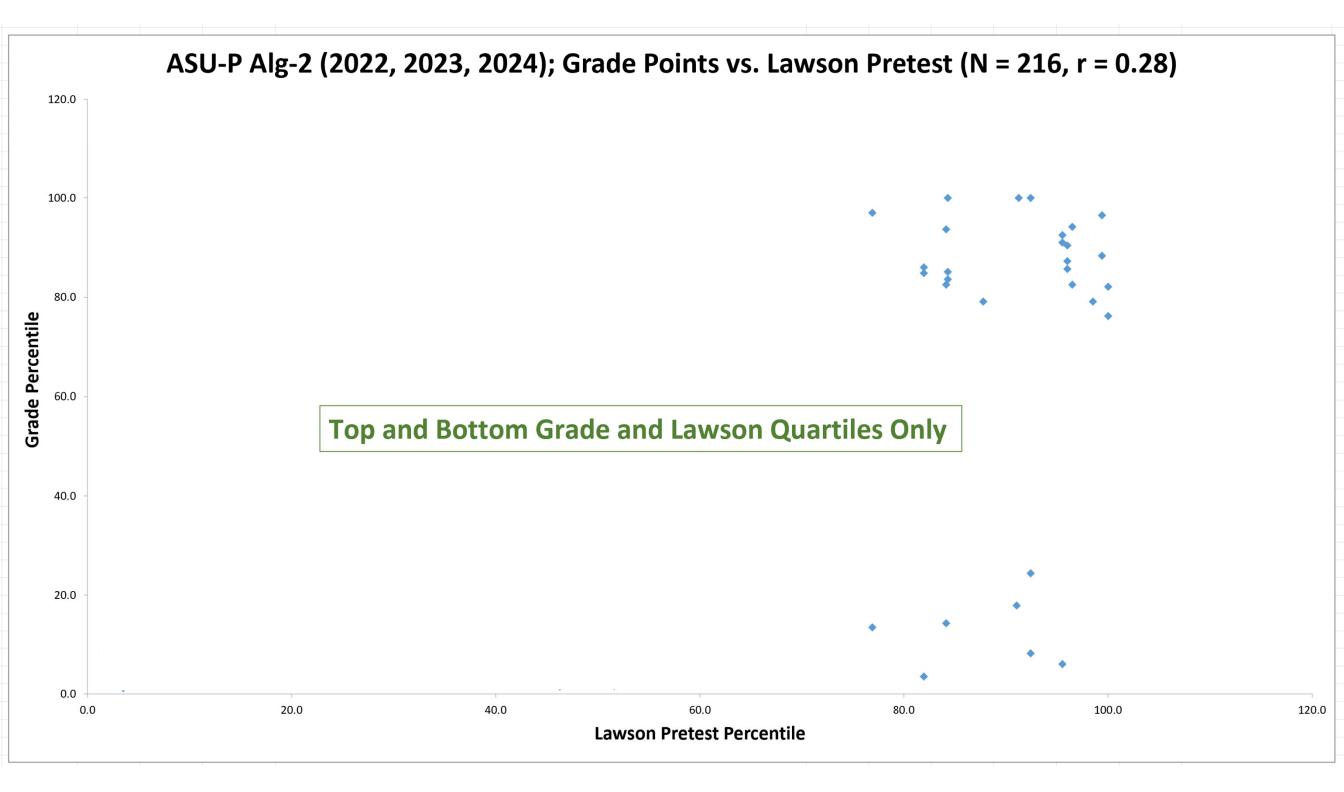
 Does using multiple predictor variables offer greater predictive power than using just one of them? Yes

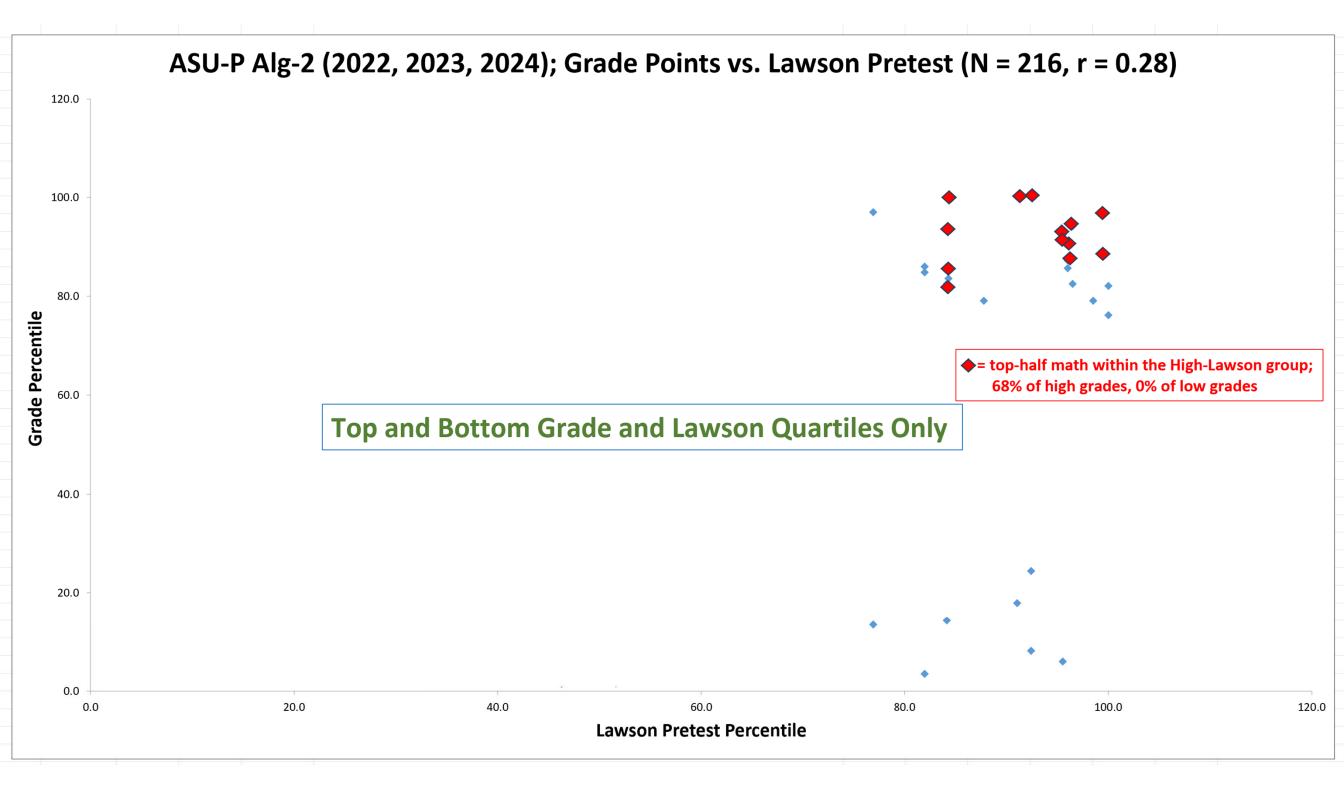




ASU-P Alg-2 (2022, 2023, 2024); Grade Points vs. Lawson Pretest (N = 216, r = 0.28)

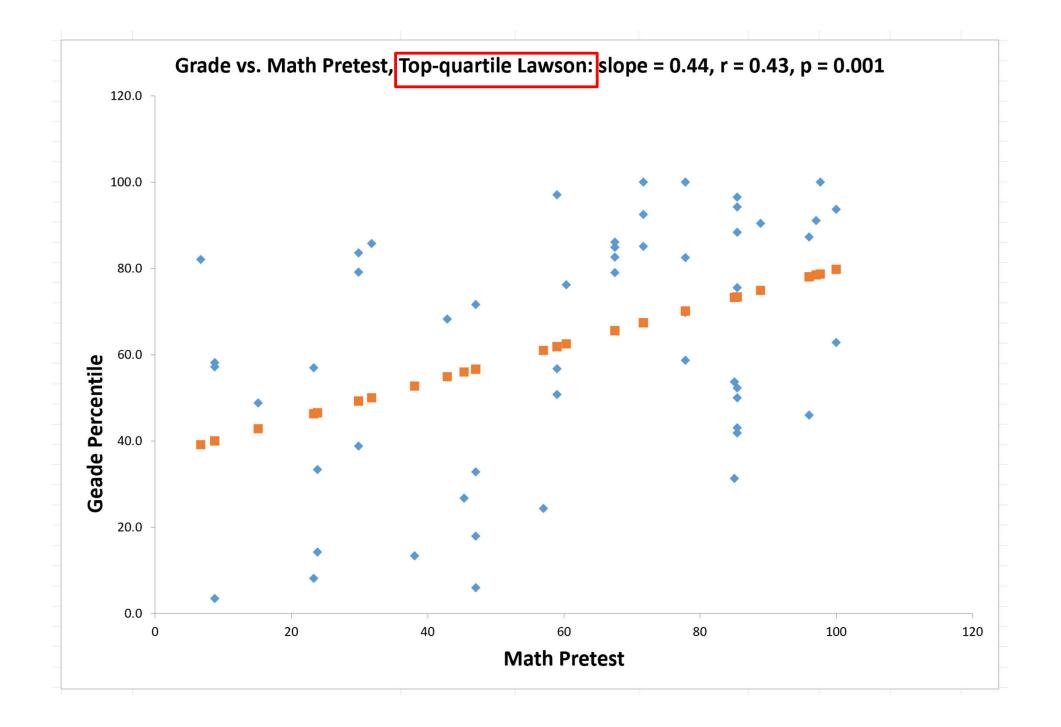


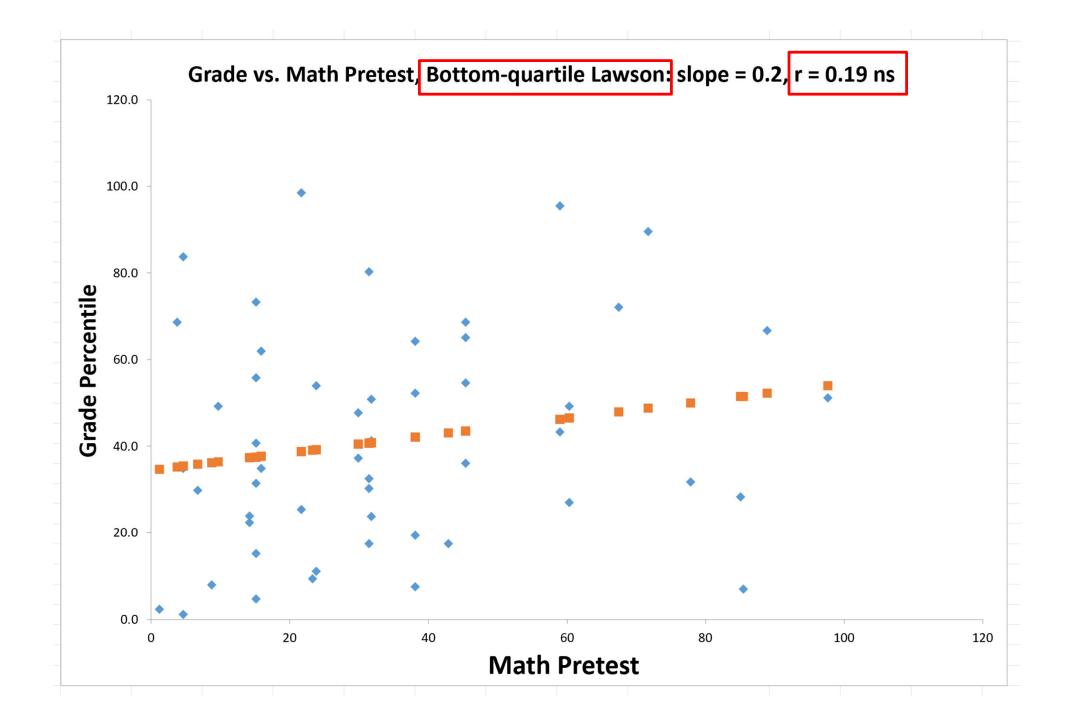




Relevant Questions

 Does better performance on one pretest indicate that another pretest is more (or less) predictive? (This would be an "interaction" effect.) *Maybe*





Important Note

- Anecdotal evidence shows:
 - Students with low pretest scores but high grades are often highly engaged and regular participants in class activity
 - High scores but low grades are often associated with missing many classes and assignments

Summary

- High and low pretest scores on diagnostic tests are consistently predictive of students' probability of attaining high or low grades
- High pre-instruction scores on a math diagnostic, a test of scientific reasoning, and the FCI are all *independently* associated with higher probability of getting high grades (and avoiding low grades)

Our results are consistent with findings reported by:

- L. Ding, PRPER **10**, 023101 (2014)]
- Salehi et al., PRPER **15**, 020114 (2019)
- Stewart et al., PRPER **17**, 010107 (2021)