Making the Best Use of Your Data

Effective Strategies for Research Partnerships and In-House Analysis

Dorothyjean Cratty
SLDS, NCES, U.S. Department of Education
dorothyjean.cratty@ed.gov

NCES Stats DC Conference, July 2012

Intended Audience for the Presentation (and Handouts):

- State and Local Education Agencies (SEAs and LEAs) conducting in-house analysis.
- SEAs and LEAs collaborating with research partners.
- Research partners and other researchers who work with administrative data in an SEA or LEA Longitudinal Data System (LDS).
- Anyone interested in conducting or interpreting administrative LDS analysis.

The Presentation Demonstrates Techniques Detailed in:

SLDS Data Use Issue Brief #2: "Forming Research Partnerships with State and Local Education Agencies"

SLDS Data Use Issue Brief #3: "Turning Administrative Data into Research-Ready Longitudinal Datasets"

SLDS Data Use Issue Brief #4: "Techniques for Analyzing Longitudinal Administrative Data"

And research conducted at Duke University (prior to NCES):

Cratty, Dorothyjean (2012). "Potential for Significant Reductions in Dropout Rates: Analysis of an Entire 3rd Grade State Cohort." *Economics of Education Review* 31(5): 644-662.

Cratty, Dorothyjean (2012). "Do 3rd Grade Math Scores Determine Students' Futures? A Statewide Student-Level Analysis of College Readiness and the Income-Achievement Gap" Forthcoming.

Descriptive Analysis with SEA or LEA Administrative Data

Simple Descriptive Reports	Descriptive Analysis	Complex or Causal Analysis
• Enrollment Counts • Course Rosters • Subgroup Proficiencies	• Correlations or Patterns • Conditional Probabilities • Predetermined Variables	• Random Control Trials • Quasi Experimental Designs • Hierarchical Linear Models
"Black and Hispanic students drop out at higher rates than white and Asian students."	"Controlling for 3rd grade math and reading scores, black and Hispanic rates are lower."	"Controlling for all else, program X reduces dropout rates by ##.##%."
Less information, but fewer assumptions about the data.	"Controlling for scores, absenteeism, and the school they attend, low-income and highly-mobile students are more likely to drop out and learning disabled students are less likely to drop out."	**More information, but more assumptions about the data.**
Primarily conducted by those who know the data best.	"Grade retention leading to .3SDs of growth is correlated with a reduction in the probability of dropping out."	**Primarily conducted by those who know the data the least.**

LDS Files Are Not Necessarily Research-Ready Datasets

LDS Administrative Data:

- Student assessment records
- Student enrollment records
- Teacher personnel files
- Course schedules
- Transcript records
- Parent information
- Federal aggregates (EDFacts)
- **Compiled from local sources**

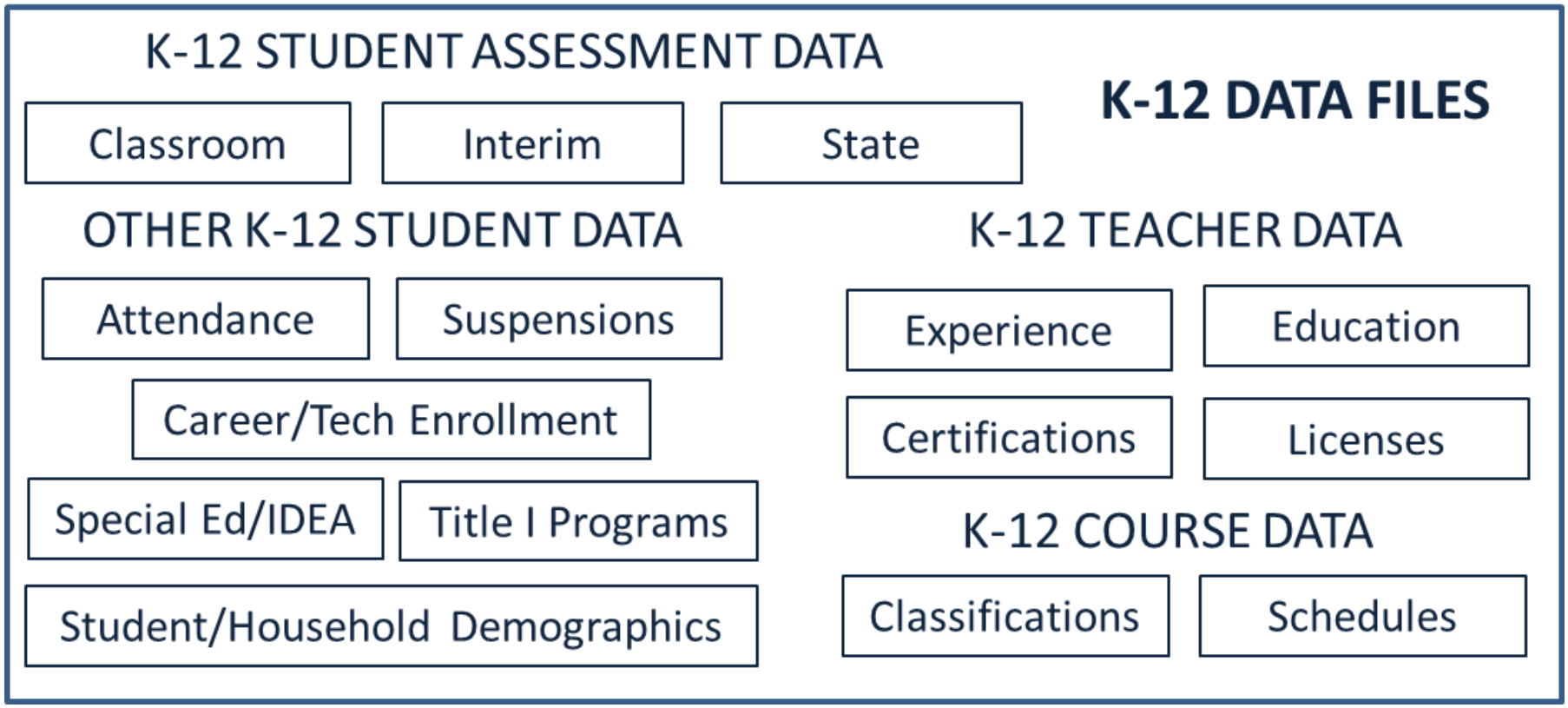
Ideal Research Data:

- Unique, encrypted student IDs
- Complete representative sample
- Data linked across files
- Data linked across years
- Student-teacher links
- Detailed course information
- Important household variables
- **Tractable data collection process**

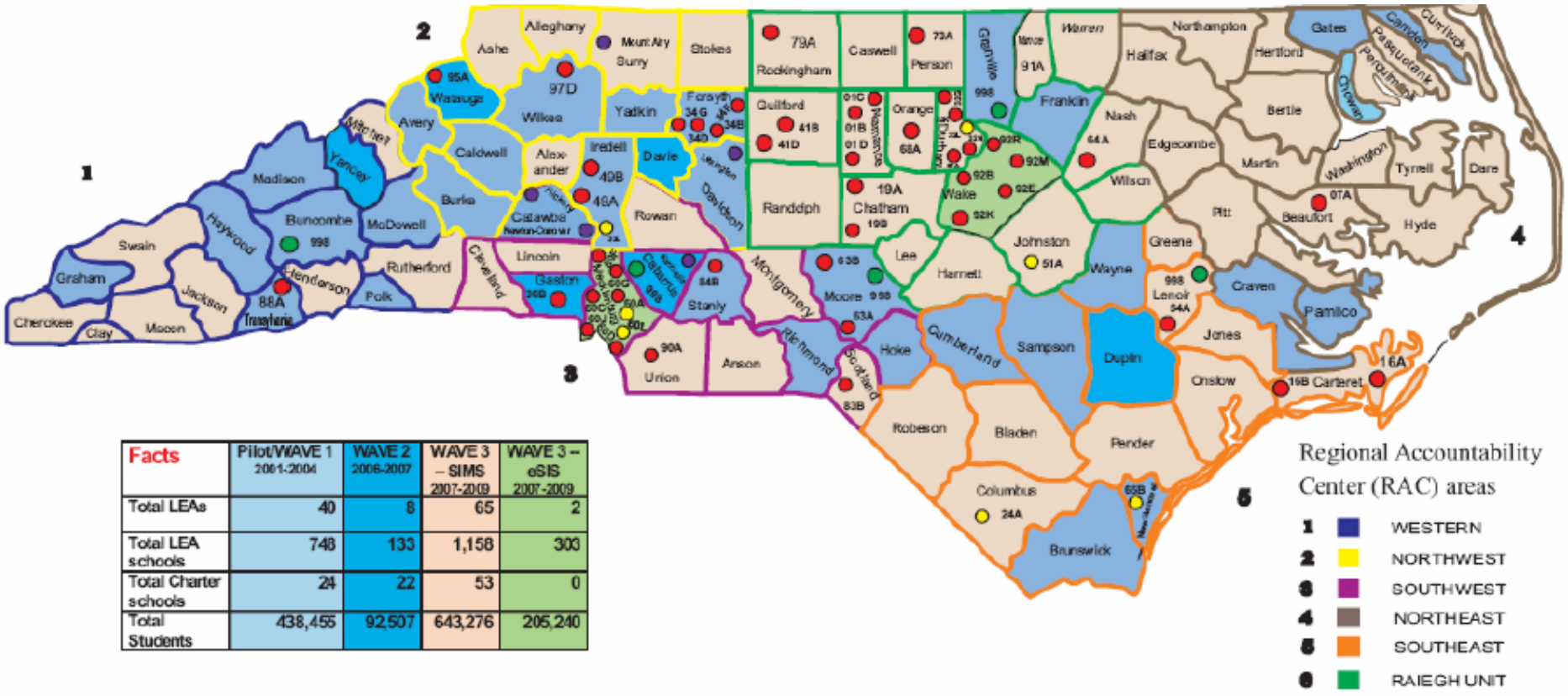
Ideal Analysis Dataset

Student	Year	Grade	School	Math Teacher	Math Class	Math Score	Math GPA	Days Absent	Days Susp.	Exit Status
1000001	2003	6	101	10006	Adv 6th	#	NA	2	0	NA
1000001	2004	7	101	10007	Pre-Algebra	#	NA	0	0	NA
1000001	2005	8	101	10008	Algebra I	#	NA	1	0	NA
1000001	2006	9	102	10009	Algebra II	#	#	3	2	NA
1000001	2007	10	102	10010	Geometry	#	#	2	0	NA
1000001	2008	11	102	10011	Pre-Calculus	NA	#	5	0	NA
1000001	2009	12	102	10012	Calculus	NA	#	4	0	Graduated
1000002	2003	6	201	20006	Std 6th	#	NA	9	0	NA
1000002	2004	7	201	20007	Std 7th	#	NA	7	2	NA
1000002	2005	8	201	20008	Math 8th	#	NA	11	0	NA
1000002	2006	9	202	20009	Algebra I	#	#	16	3	NA
1000002	2007	9	203	20010	Algebra I	#	#	22	5	NA
1000002	2008	10	203	20011	none	NA	#	28	8	Dropped Out
1000003	2003	6	301	30006	Std 6th	#	NA	3	.	NA
1000003	2004	7	302	30007	Std 7th	#	NA	1	2	NA
1000003	2005	8	302	30008	Math 8th	#	NA	0	0	NA
1000003	2006	9	302	30009	Algebra I	#	#	7	3	NA
1000003	2007	10	303	30010	Algebra I	.	#	5	0	NA
1000003	2008	11	303	30011	Algebra II	#	#	9	4	NA
1000003	2009	12	303	30012	Geometry	#	#	2	0	Graduated

Longitudinal Data Systems Likely Comprise Separate Files



The Local Data Collection Process May Vary Widely



Facts	Pilot/WAVE 1 2001-2004	WAVE 2 2006-2007	WAVE 3 – SIMS 2007-2009	WAVE 3 – eSIS 2007-2009
Total LEAs	40	8	65	2
Total LEA schools	748	133	1,156	303
Total Charter schools	24	22	53	0
Total Students	438,455	92,507	643,276	205,240

Pilot & WAVE 1 2001–2004

WAVE 2 2006-2007

WAVE 3 — SIMS 2007-2009

- Existing Charter Schools
- City Schools
- New Charter Schools
- Agency Schools

WAVE 3 --- eSIS 2007-2009

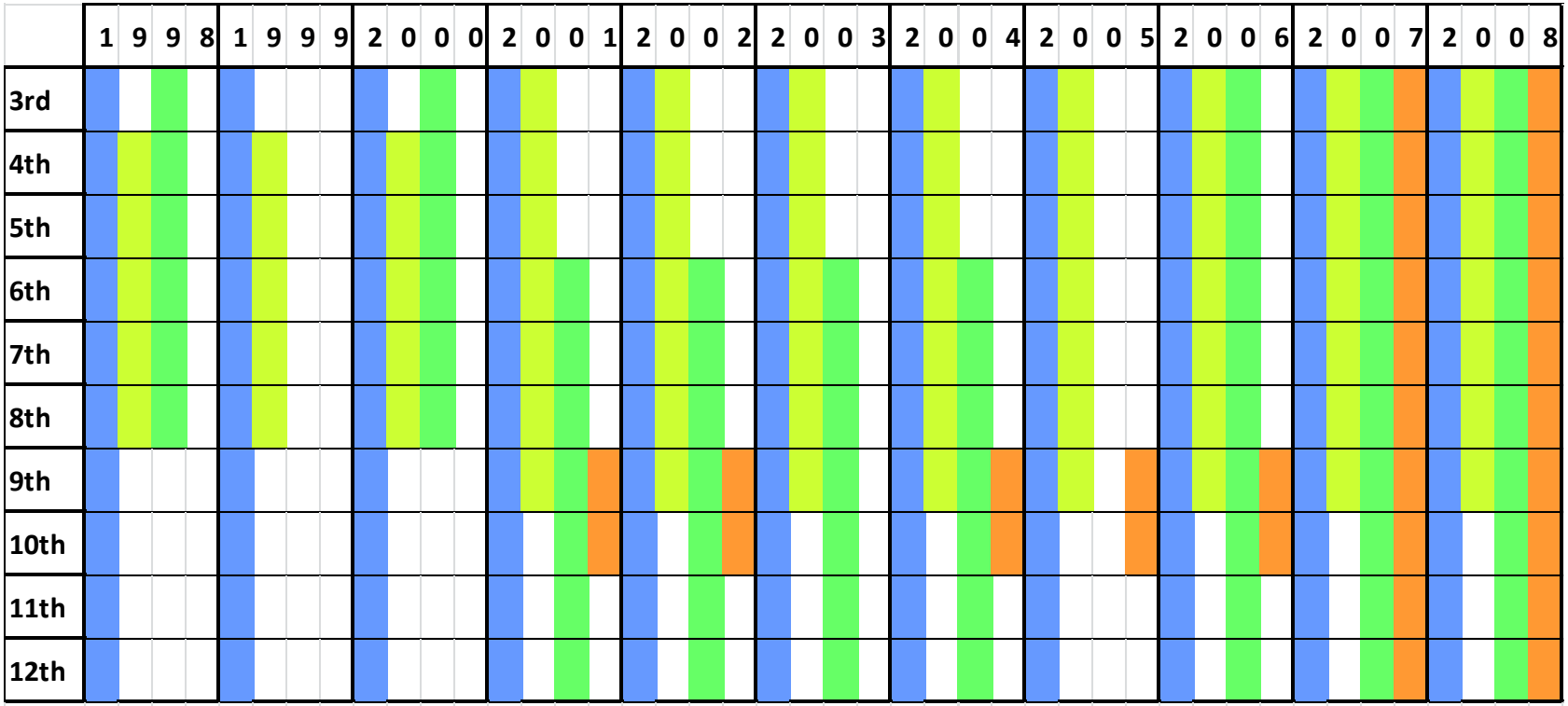
Wake: 148 schools
Mecklenburg: 155 schools
Total students: 205,240

Source: North Carolina Window of Information on Student Education (NCWISE) systems deployment map.

Suggestions for Conducting Research with LDS Data

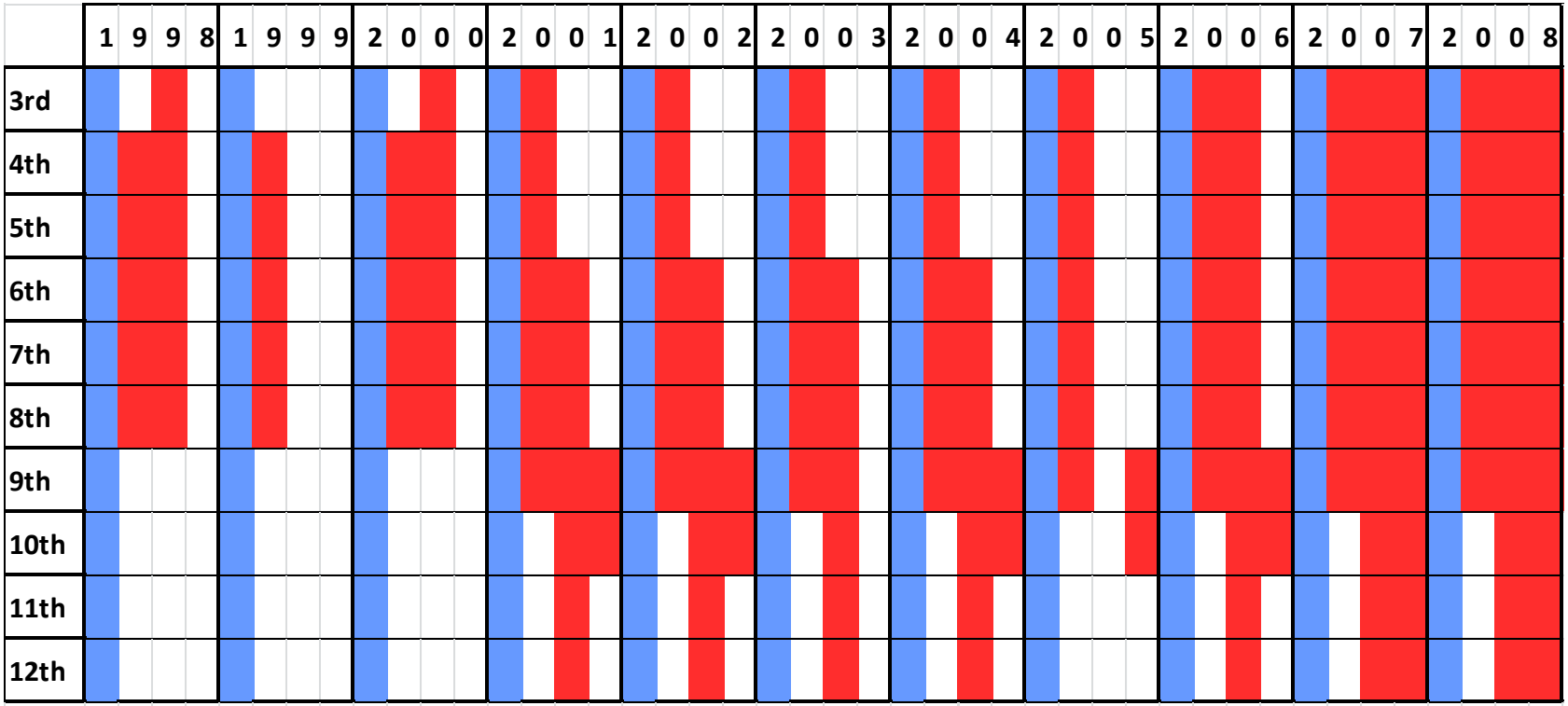
- The best LDS research combines solid methods with knowledge of the data collection process, including variation in local education policies.
- Limiting analysis to a subset of readily available files can lead to underuse and/or misuse of important data for studying education.
- It is important to know why and how each record was collected; make use of meta/para-data on SEA, LEA, and federal reporting websites.
- Use related data sources to triangulate missing or conflicting data points and to check assumptions, and be explicit about all data decisions.

Examples Using North Carolina Student-level Data Files



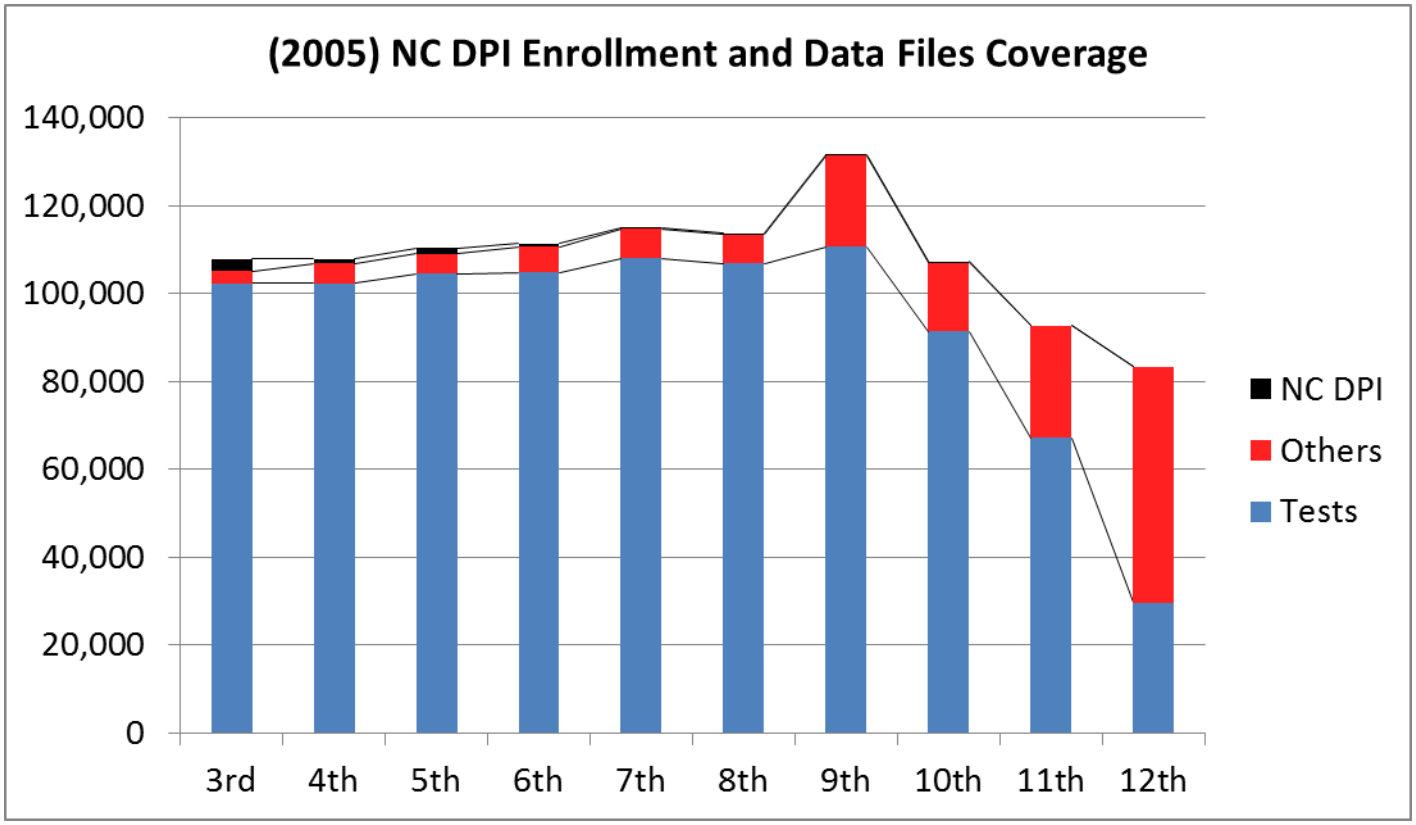
TESTS	CHARACTERISTICS	DISCIPLINE	COURSES
Scores	Demographics (Age, Race, Gender)	Suspension (Some Years)	CTE (2001-2006)
Proctors	SES (Parents' Education, School Lunch)	Absenteeism (Some Years)	All (2007-2008)
Classmates	Special Education Services (LD, Gifted, etc.)	Tardys (Some Years)	Teachers (2007-2008)

Most Research with NC Data Uses Test Files Primarily



TESTS	CHARACTERISTICS	DISCIPLINE	COURSES
Scores	Demographics (Age, Race, Gender)	**Suspension** (Some Years)	CTE (2001-2006)
Proctors	SES (Parents' Education, School Lunch)	Absenteeism (Some Years)	All (2007-2008)
Classmates	Special Education Services (LD, Gifted, etc.)	Tardys (Some Years)	Teachers (2007-2008)

Test Files are Good for 3rd-8th Tests and Demographics



- Test Records Cover 95% of 3rd-8th Grade Students and High School Tests
- Tests Plus All Administrative Records Equal Total NC Student Enrollment
- Administrative Files Have Courses, Teachers, Programs, Suspensions, etc.

Descriptives Using Both Test and Administrative Data

	Total Cohort	Graduates	Dropouts	Difference (Grad-Drop)
Percent of All 1998 NC 3rd Graders	100.0%	80.7%	19.3%	
3rd Grade EOG Math z-score	0.06	0.18	-0.46	0.65***
3rd Grade EOG Reading z-score	0.06	0.18	-0.49	0.67***
Mean 3rd to 8th Math z-score	0.04	0.19	-0.58	0.76***
Mean 3rd to 8th Reading z-score	0.02	0.17	-0.60	0.77***
Growth in 3rd to 8th Math z-scores	0.005	0.012	-0.024	0.035***
Growth in 3rd to 8th Reading z-scores	0.007	0.009	-0.005	0.014***
Mean GPA	2.66	2.86	1.64	1.23***
				(std errs)

***p<0.001. The standard errors from t-tests of the difference in means are statistically significant at the 1% level.
Source: "Potential for Significant Reductions in Dropout Rates" (Cratty, 2012).

Correlates of High School Dropout Combining All the Data

Determinants as of:	3rd Grade	5th Grade	8th Grade	9th Grade
Initial EOG Math z-score	-0.288***	-0.339***	-0.440***	-0.318***
Average EOG Math Growth		-0.695***	-2.961***	-2.378***
Initial EOG Reading z-score	-0.259***	-0.269***	-0.218***	-0.137***
Average EOG Reading Growth		-0.441***	-1.066***	-0.647***
School Moves: One (vs. None)		0.207***	0.262***	0.210***
School Moves: Mult (vs. One)		0.311***	0.355***	0.325***
Days Abs: 8-14 (vs. 0-7)		0.555***	0.505***	0.448***
Days Abs: 15-21 (vs. 8-14)		0.382***	0.356***	0.296***
Days Abs: Over 21 (vs. 15-21)		0.369***	0.319***	0.329***
Retention: 3rd-5th (vs. None)		1.030***	1.134***	1.168***
Retention: 6th-8th (vs. None)			1.609***	1.539***
Retention: 3rd-5th and 6th-8th			3.778***	3.428***
Retention 9th: One (vs. None)				1.913***
Retention 9th: Mult (vs. One)				1.570***
Out Suspension: One (vs. None)			0.671***	0.575***
Out Suspension: Mult (vs. One)			0.589***	0.368***
Algebra I by 8th Grade			-0.507***	-0.451***
Middle School in 6th Grade			n.s.	n.s.

A subset of the logit coefficients from the probability models.

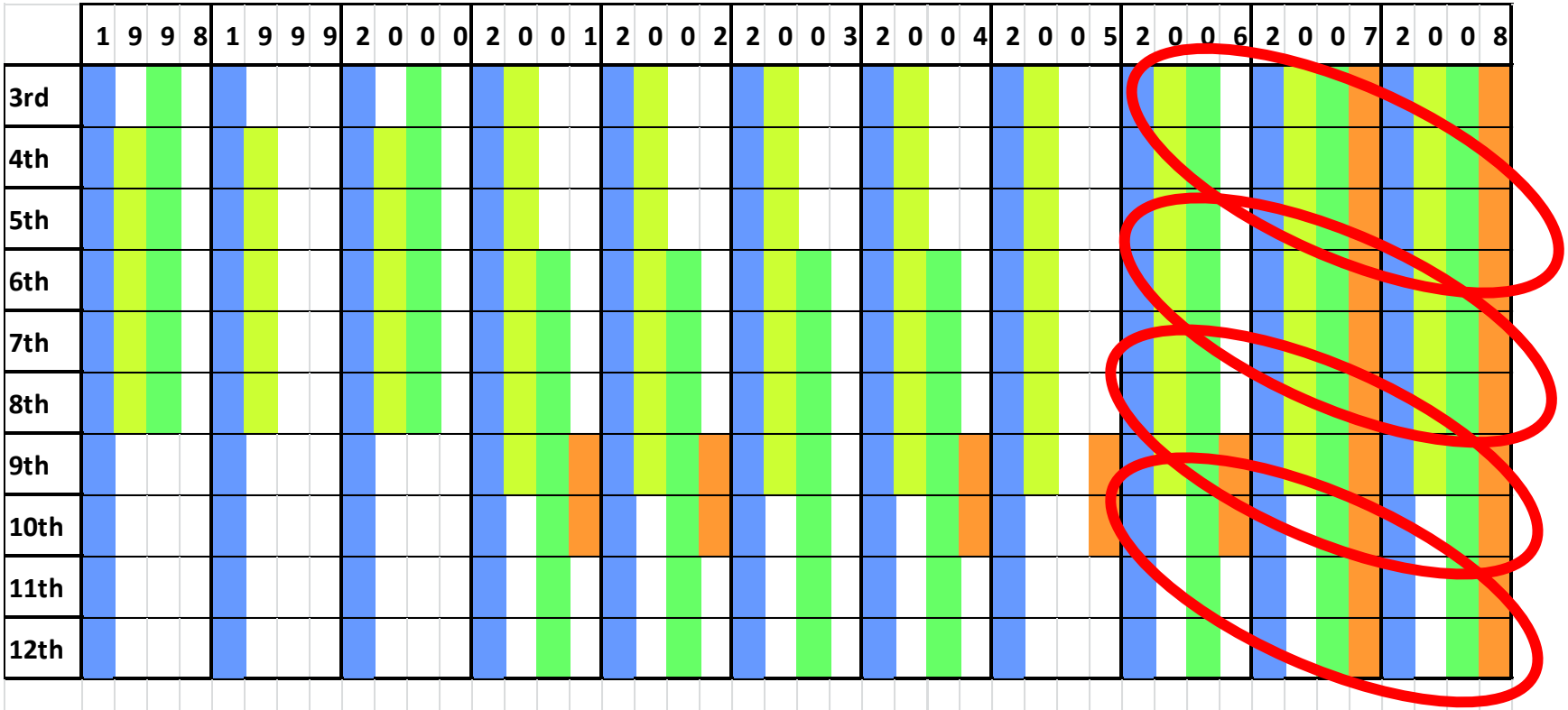
Dropout Simulations Incorporating Correlations in the Data

Table 6: Reductions in Dropout Rates Resulting from Various Simulations Using 8th Grade Logits.

	Percent Reduction in Pr(D) for At-Risk Students (without Intermediate Effect)	Percent Reduction in Pr(D) for At-Risk Students (with Intermediate Effect)	Total Reduction in Dropouts for Entire Cohort (Percent, Rate, & Number)
Adopt AIG Model for 16,000 3rd Graders with Highest Pr(D) and Keep Existing Program	19.00%	43.82%	24.99% 18.3 to 13.7 3,024 fewer
Adopt AIG Model for 16,712 3rd Graders Below Proficiency and Keep Existing Program	20.61%	47.08%	20.61% 18.3 to 14.52 2,494 fewer
Adopt AIG Model for 17,293 Barely Proficient 3rd Graders and Keep Existing Program	24.49%	48.06%	14.38% 18.3 to 15.7 1,740 fewer
Provide AIG Model to 8,000 3rd Graders with Highest Pr(D) and 8,000 with Highest Ability	16.16%	39.99%	12.47% 18.3 to 16.0 1,509 fewer

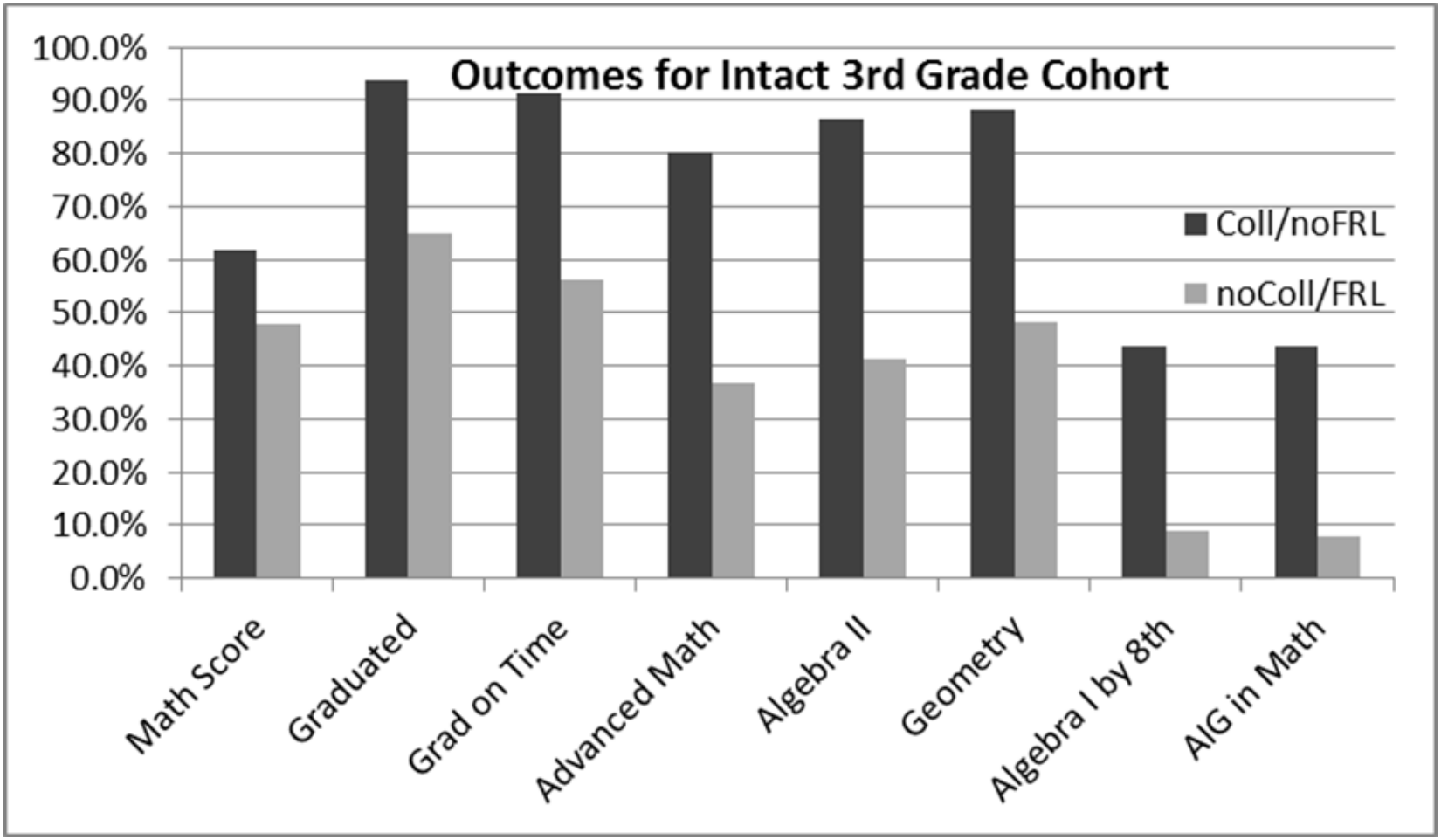
The correlation of 0.10SD annual test score growth with probabilities of dropping out of high school for different subsets of students.

College Readiness Analysis Adds in Recent Course Data



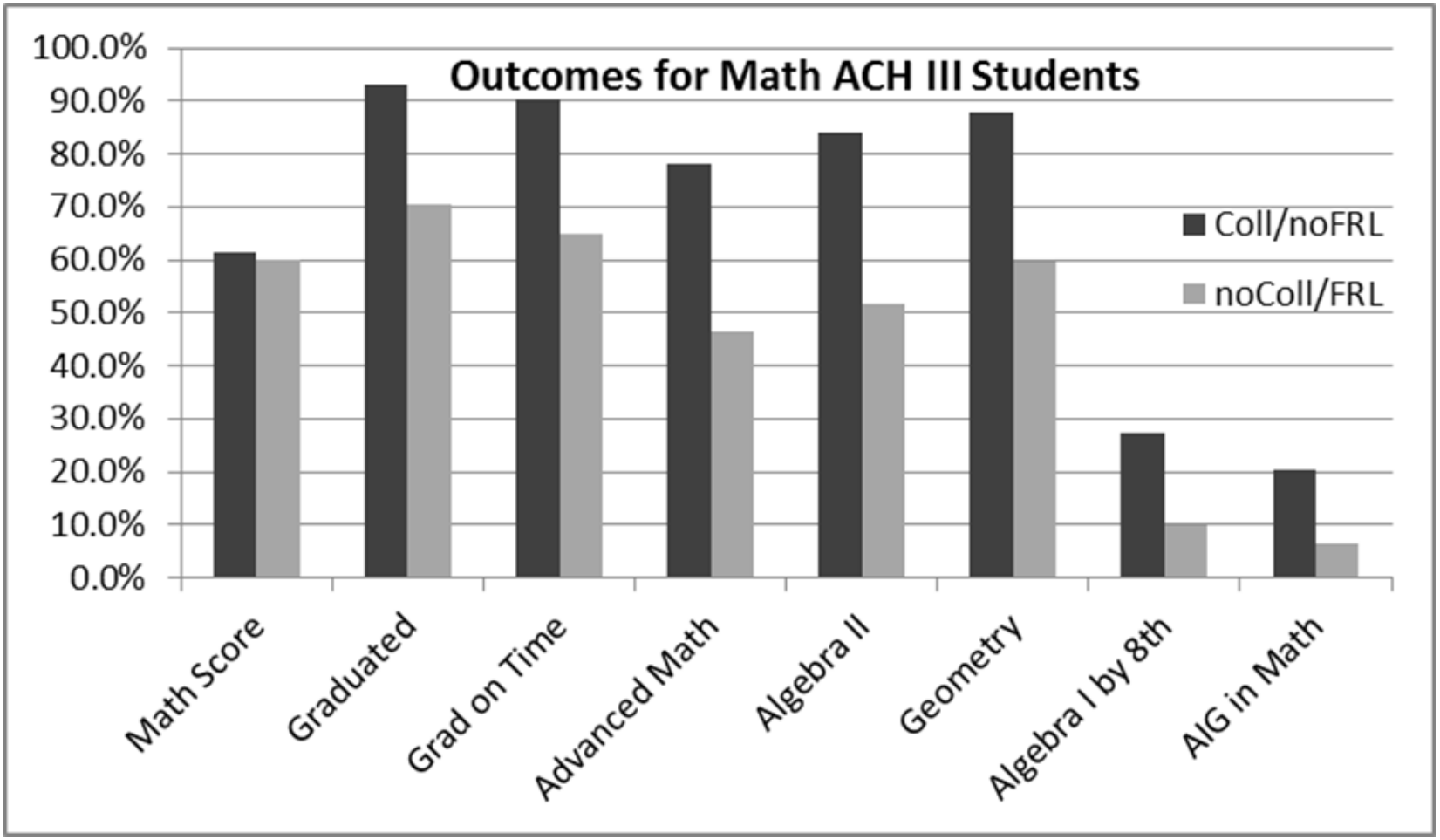
TESTS	CHARACTERISTICS	DISCIPLINE	COURSES
Scores	Demographics (Age, Race, Gender)	Suspension (Some Years)	CTE (2001-2006)
Proctors	SES (Parents' Education, School Lunch)	Absenteeism (Some Years)	All (2007-2008)
Classmates	Special Education Services (LD, Gifted, etc.)	Tardys (Some Years)	Teachers (2007-2008)

College Readiness Outcomes for 1998 3rd Graders by SES



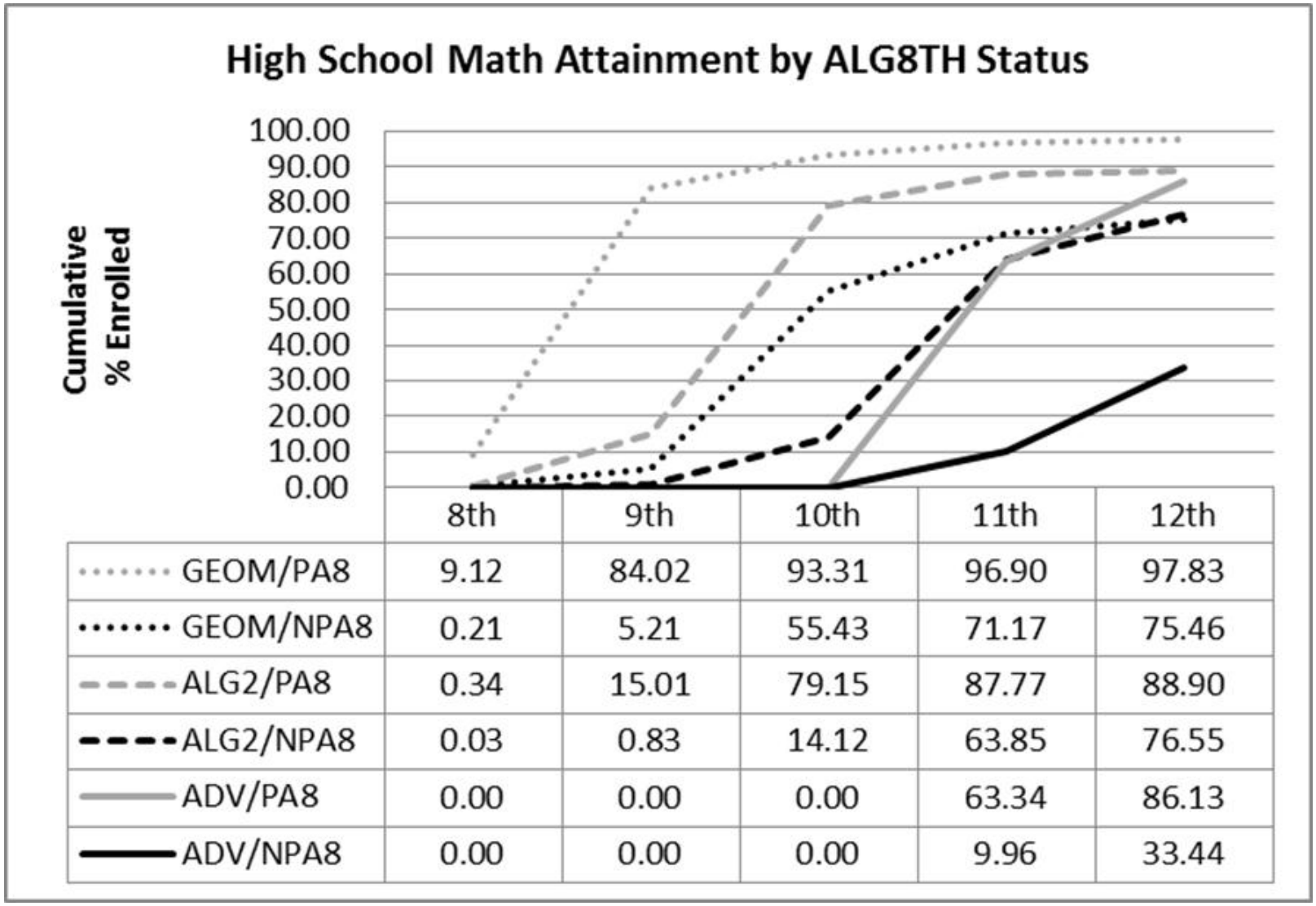
Source: "Do 3rd Grade Math Scores Determine Students' Futures?" (Cratty, 2012). Note: SES uses parents' education and lunch eligibility

Outcomes for Math Proficient 3rd Graders by SES



Source: "Do 3rd Grade Math Scores Determine Students' Futures?" (Cratty, 2012). Note: SES uses parents' education and lunch eligibility

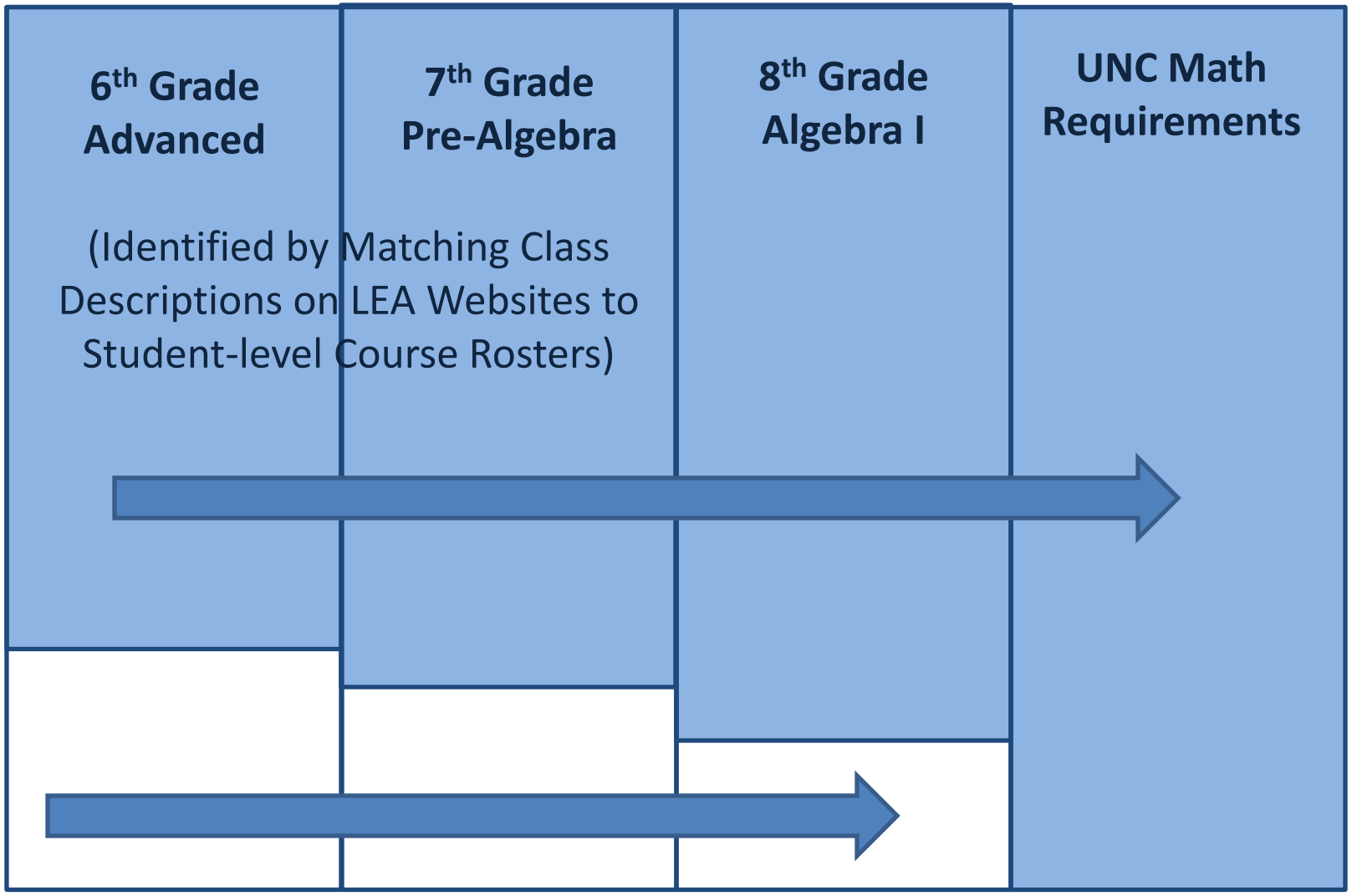
College Prep Math Course Trajectories by 8th Grade Algebra



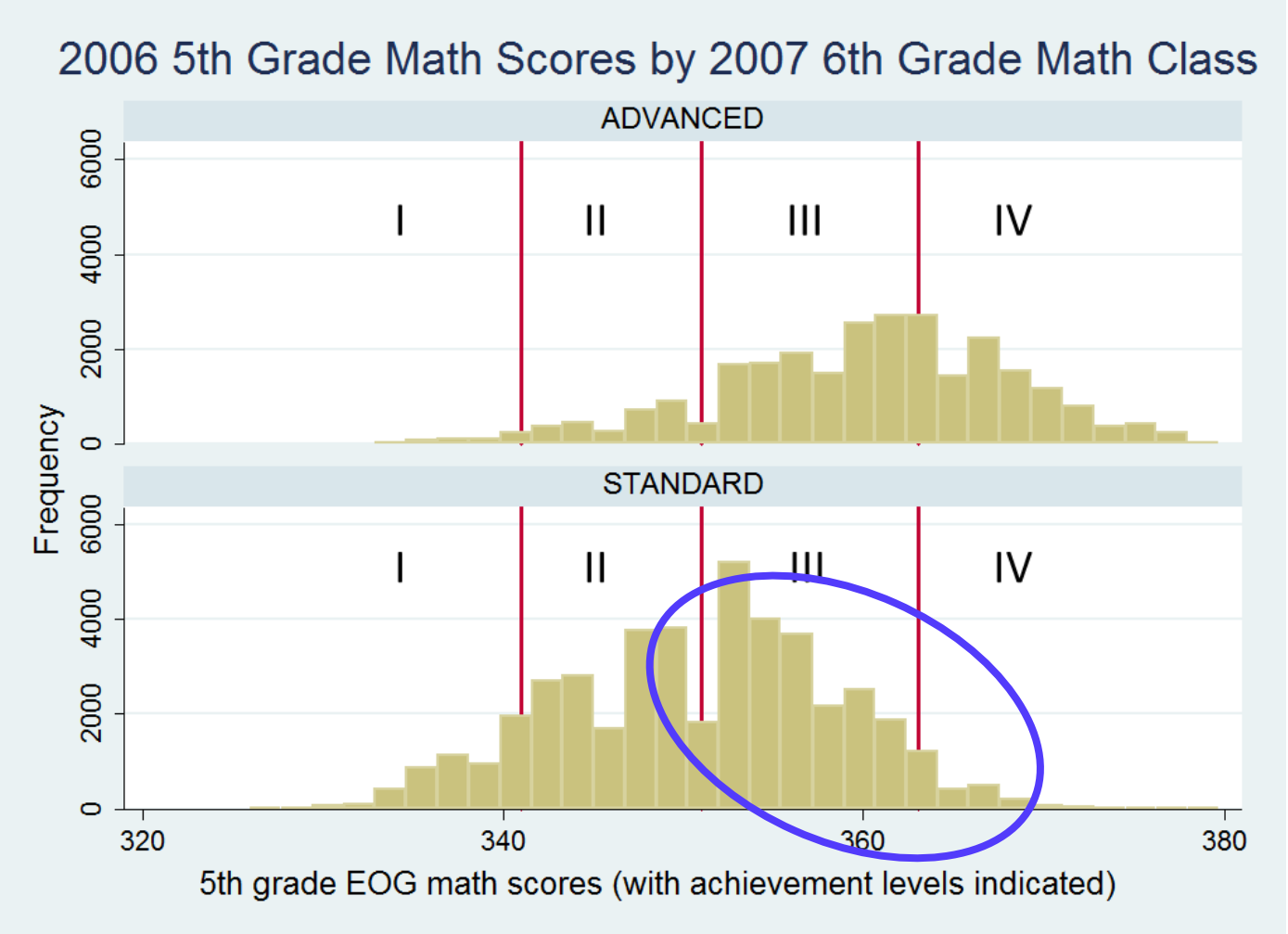
	8th	9th	10th	11th	12th
GEOM/PA8	9.12	84.02	93.31	96.90	97.83
GEOM/NPA8	0.21	5.21	55.43	71.17	75.46
ALG2/PA8	0.34	15.01	79.15	87.77	88.90
ALG2/NPA8	0.03	0.83	14.12	63.85	76.55
ADV/PA8	0.00	0.00	0.00	63.34	86.13
ADV/NPA8	0.00	0.00	0.00	9.96	33.44

Source: "Do 3rd Grade Math Scores Determine Students' Futures?" (Cratty, 2012). Note: NPA8 for "not passed Algebra I by 8th grade."

Important College Readiness Course Data is Not Readymade



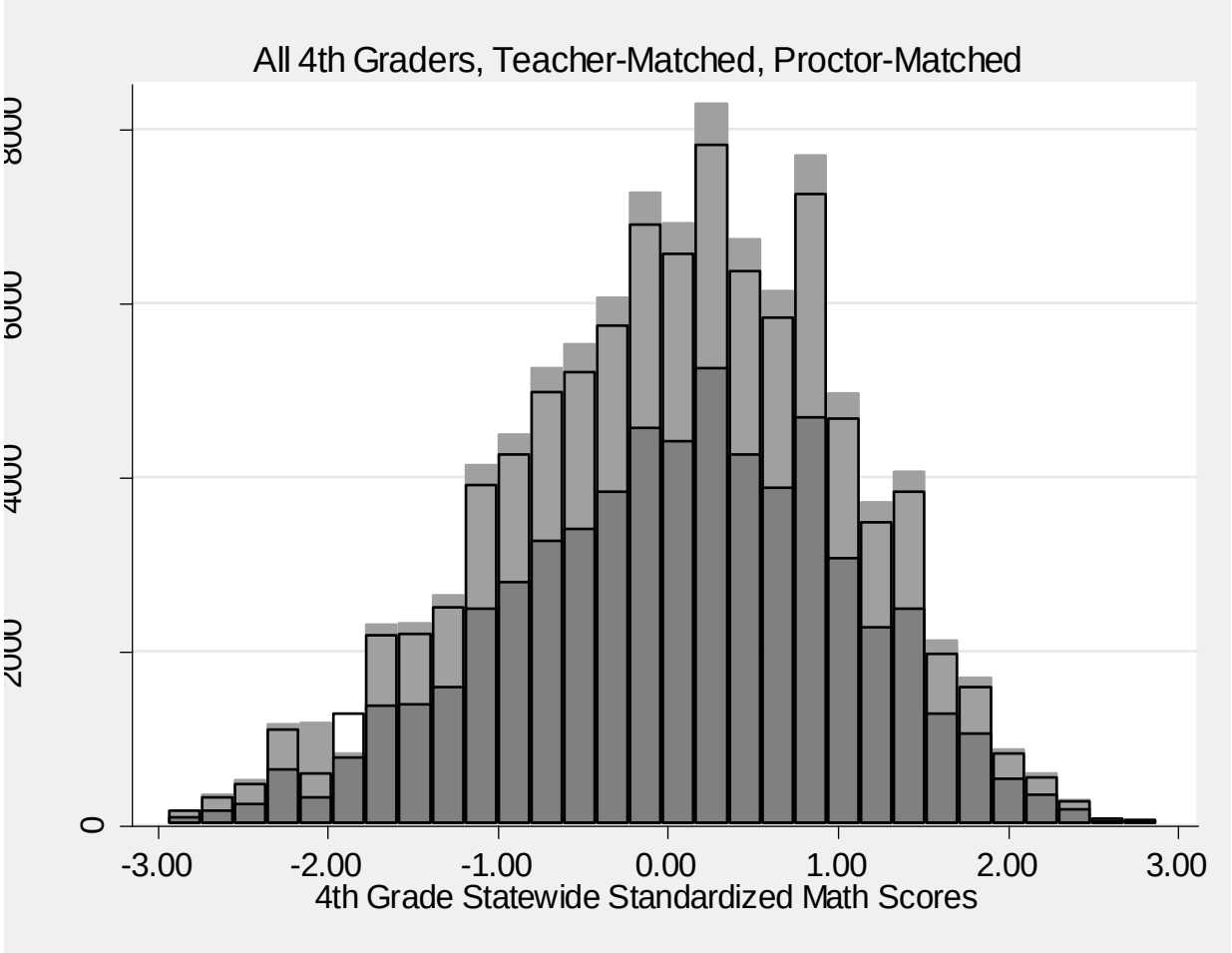
Combining Test, Course, Absenteeism, and Suspension Data



Source: "Do 3rd Grade Math Scores Determine Students' Futures?" (Cratty, 2012).

Controlling for 5th grade math scores and absenteeism, proficient 5th grade students placed in standard math had higher middle grade suspensions & unexcused absences.

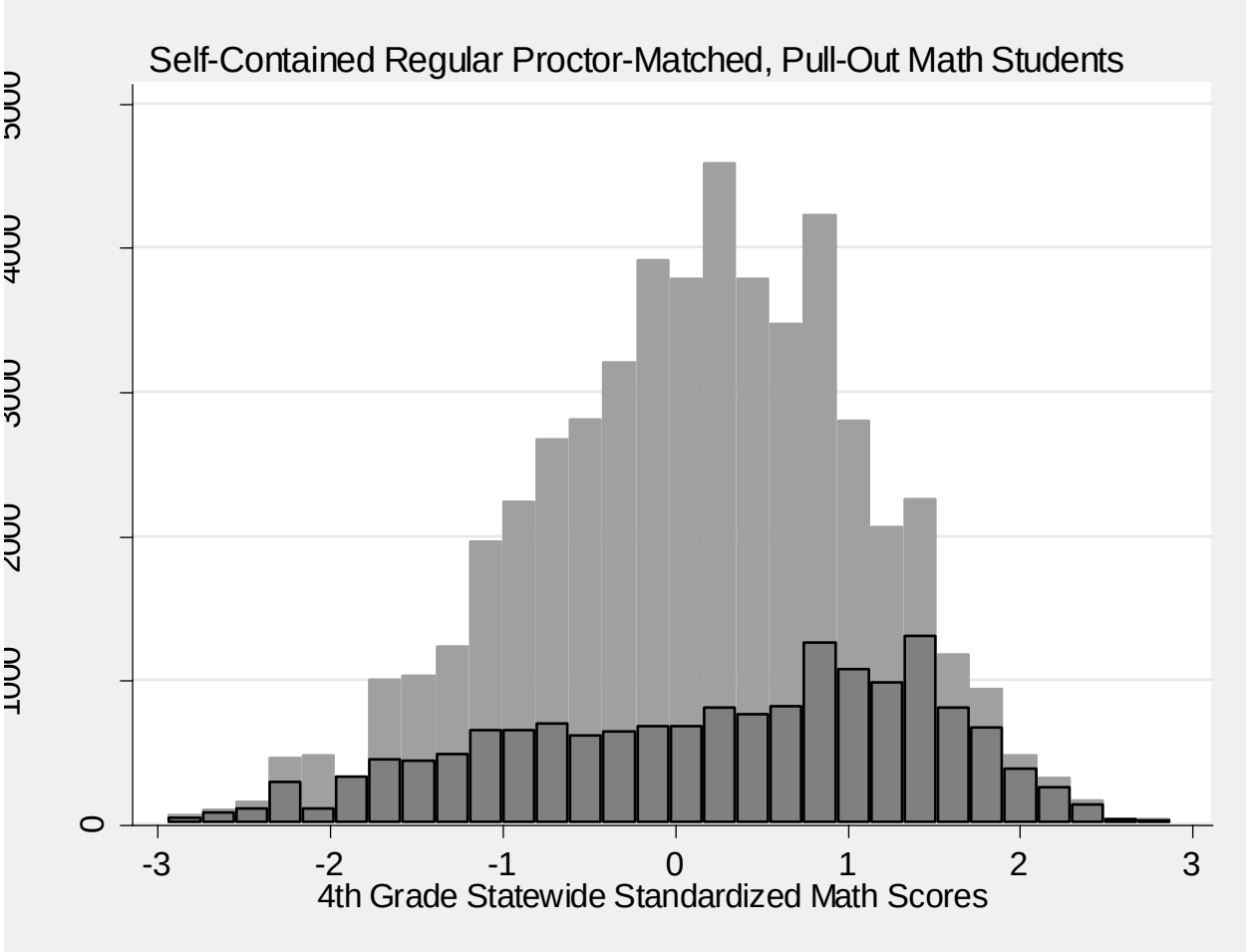
Combining Test, Classroom, Teacher, and Program Data



Source: "Do 3rd Grade Math Scores Determine Students' Futures?" (Cratty, 2012).

95% of 1999 4th grade teachers matched to students (prior to a link) by using all the information in the classroom data files vs. relying on the proctor variable in test files.

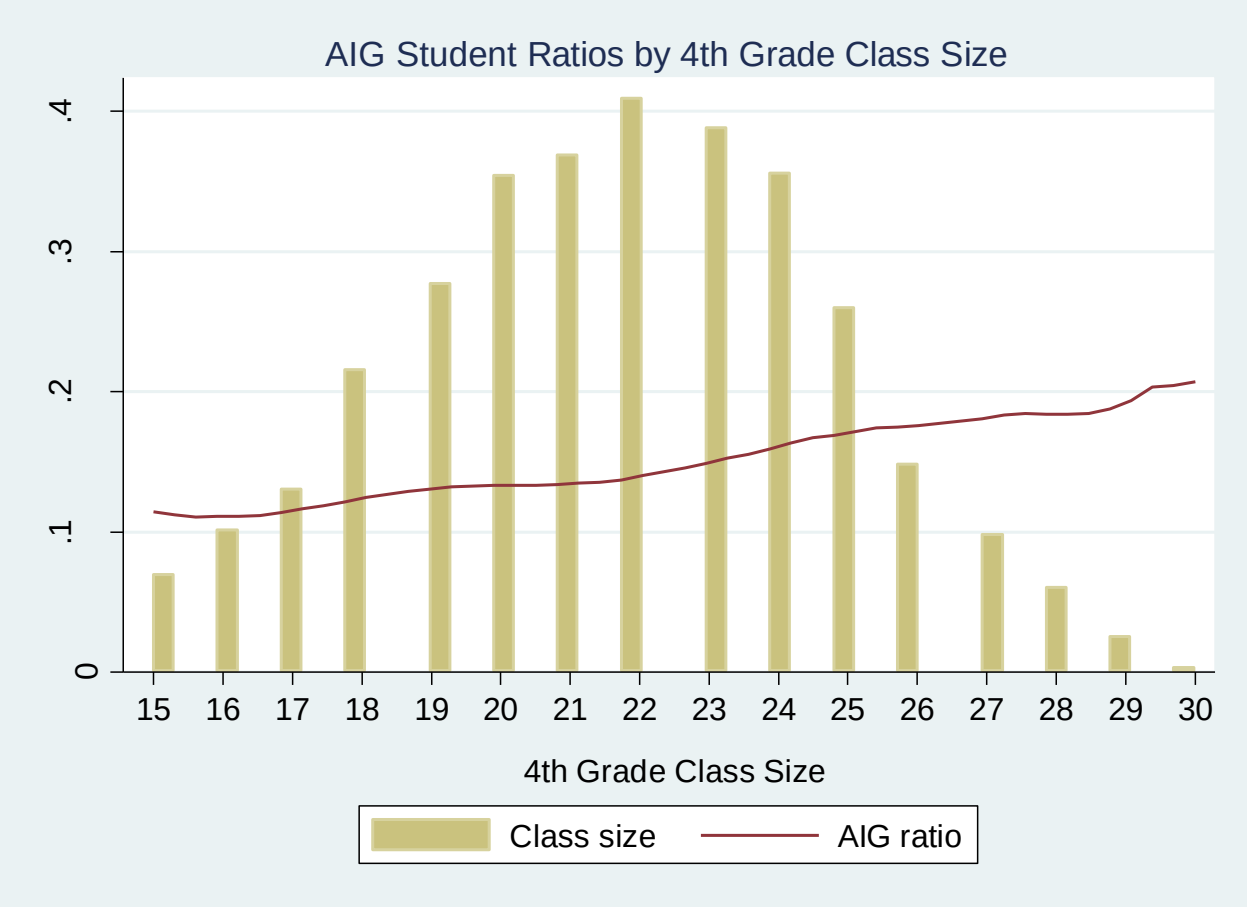
Data Also Shows Extent of Non "Self-Contained" Teaching



Source: "Do 3rd Grade Math Scores Determine Students' Futures?" (Cratty, 2012).

One out of four 4th graders are taught math by someone other than a single "self-contained" classroom teacher. Many are students in special enrichment programs.

Understanding of the Data Can Inform Class Size Studies



Source: "Do 3rd Grade Math Scores Determine Students' Futures?" (Cratty, 2012).

Larger classes have higher shares of students in the special enrichment programs. And this is true within schools and controlling for math and reading test scores.

Data Can Be Used to Replace Assumptions with Information

- Any analysis method can benefit from incorporating important information about the data and the educational context.
- State and Local Education Agency staff have the most information about the data and the context.
- For example, SEA/LEA teacher-effects models make better use of the data than most advanced academic research on the same topic using the same datasets.
- Descriptive analysis can provide valuable information about education with fewer assumptions.
- The main technique for conducting descriptive analysis is really just incorporating as much available information about the data as possible.

Conclusion

Technical assistance for LDS data analysis is available through various NCES programs, resources, and collaborations:

- SLDS Grants Program: TA requests (for all states, not just grantees), publications, webinars, working groups, and conferences (see our website).
- CEDS: NCES assistance with using the Common Education Data Standards Elements and Policy Use Cases to frame LDS research questions spanning P-20W+.
- RELs: NCES collaboration with NCEE program for Regional Education Lab research alliances with states and districts.
- Contact us for assistance: **dorothyjean.cratty@ed.gov**