

When does a “Growth model” act
the same as a “Status model”?:
Lessons learned from some empirical
growth model comparisons

*Jenn Dunn, Xin Wei, Jessica Allen, Damian
Betebenner, & Brian Gong*

Center for Assessment

Presentation to the CCSSO Accountability

Systems & Reporting (ASR) SCASS

September 25, 2007 Nashua, NH

Background

- NCLB (and many states' accountability systems prior to NCLB) focus on Status
- Many states and school-level educators interested in Growth as alternative to Status for school accountability
- Many “growth models” proposed, some for NCLB
- USED & its Peer Reviewer establish criteria for USED Growth Model Pilot, and eventually approve 9 states (and not approve at least seven states' proposals)



Background - continued

- States approved for USED GMP (Growth Model Pilot) begin reporting results in 2006; now reporting results for 2007
- Initial results are that very few schools made Growth who did not also make Status for AYP (less than 4 per 1000 schools, or 0.4 percent in 2006 for TN & NC)
- More in 2007: DE (around 4%; 6 schools); IA (around 9%; 110+ schools); TN (around 1%; 19 schools); OH (TBA – Nov.?)



Questions

- Do these results (high overlap between schools meeting Status and Growth for AYP) hold for all the states?
- Why these results? Is it due to the models, parameters, the state data, some interaction, or...?
- What are some good design principles for growth models for accountability?
- What growth models are poor for accountability but good for other uses?



Approach for Analysis

- Used *the same state data set* and applied Status and different growth models
 - This allows us to compare results, focusing on the models because we control for the data set
 - Did two studies, using two data sets and two combinations of models
- Study 1: Compared Status, Safe Harbor, and seven USED-approved GMP growth models
- Study 2: Compared Status and four non-approved growth models (some “approved model” but different parameters)



Some Quick Reminders

- Assessment and Accountability are different
 - Assessment = measurement of what is (& some interpretation)
 - Accountability = evaluation in comparison to what “ought to be/good enough” and some consequences
- There are many possibilities for accountability of what to measure, how to measure (general and specific), and “good enough”



Quick Reminders - continued

- USED GMP defines “good enough” student growth and “good enough” school performance very specifically
 - Student growth is anchored in proficient (must be “sufficient growth to be proficient” within set time)
 - Time is limited to five years maximum for students
 - School performance is calculated by “counting students who are proficient/on track to be proficient”
 - “Enough students proficient” for AYP is expressed as an AMO for growth
 - AMO for Growth usually calculated like AMO for Status



School Quality

	How good is this school? (Status)	Is it getting better? (Improvement)
Achievement (AYP)	(A) What is the achievement level of the students in this school?	(B) Is the achievement level of this changing over time?
Effectiveness (Typical Growth)	(C) Is this an effective school? Given the achievement level of students when they enter, how much do they learn or develop while they are in the school?	(D) Is this school becoming <i>more</i> effective? How much more, or less, are the students learning this year than they did the year before?
Growth towards proficiency (AYP Growth)	(E) Are students on track to proficiency?	(F) Is the school increasing the number of students who are on track to being proficient?

Adapted from Carlson, D. (2002). Focusing State Educational Accountability Systems: Four Methods of Judging School Quality and Progress.



Findings of Studies 1 & 2

- Do the NCLB-approved growth-pilot models classify schools differently than Status?
 - NO – largely the same

- Are there some growth models whose results do differ from Status?
 - YES – “non-approved”

- What is going on?



Study 1

- Used two years' real data from a state
- Applied growth models from seven states:
AK, AR, AZ, DE, FL, IA, NC
- Calculated Status, Safe Harbor, and Growth
- Checked for agreement in accountability judgment (decision consistency) between Status, Safe Harbor, and each of seven Growth models



Study 1 Results - Correlations

	AR	AZ	IA	FL	NC	DE
AK	.99	.92	.87	.93	.81	.76
AR		.90	.89	.92	.82	.76
AZ			.79	.85	.73	.73
IA				.83	.88	.77
FL					.76	.67
NC						.70



Status & Growth School Counts

In General:

- As the AMO increases
 - Fewer schools meet status
 - More schools meet growth
- The growth models are classifying similar numbers of schools

Status		Y	N	Y	N	Met
Growth		Y	N	N	Y	AYP
58% proficient	AK	116	19	0	5	121
	AR	116	19	0	5	121
	AZ	116	20	0	4	120
	IA	116	15	0	9	125
	FL	116	18	0	6	122
	NC	116	11	0	13	129
	DE	116	22	0	2	127
72% proficient	AK	78	43	0	19	97
	AR	78	42	0	20	98
	AZ	78	44	0	18	96
	IA	78	42	0	20	98
	FL	78	35	0	27	105
	NC	78	35	0	27	105
	DE	78	57	0	5	103
80% proficient	AK	47	71	0	22	69
	AR	47	72	0	21	68
	AZ	47	77	0	16	63
	IA	47	70	0	23	70
	FL	47	66	0	27	74
	NC	47	63	0	30	77
	DE	47	88	0	5	72



Growth & Confidence Intervals

In General:

- The CI & GM tend to “help” the same schools
- The CI tends to “help” more schools

CI Growth		95% CI			99% CI		
		Y	Y	N	Y	Y	N
		Y	N	Y	Y	N	Y
72% Proficient	AK	14	2	5	15	8	4
	AR	13	3	7	16	7	4
	AZ	12	4	6	15	8	3
	IA	12	4	8	15	8	5
	FL	15	1	12	21	2	6
	NC	13	3	14	17	6	10
	DE	3	13	2	5	18	0
80% Proficient	AK	16	3	6	21	6	1
	AR	16	3	5	20	7	1
	AZ	13	6	3	16	11	0
	IA	14	5	9	18	9	5
	FL	18	1	9	24	3	3
	NC	16	3	14	20	7	10
	DE	3	16	2	4	23	1



Growth Models & Safe Harbor

		Status model			
SH		Y	N	Y	N
GM		Y	N	N	Y
58% Proficient	AK	1	19	0	4
	AR	1	19	0	4
	AZ	1	20	0	3
	IA	1	15	0	8
	FL	1	18	0	5
	DE	0	21	1	2
72% Proficient	AK	2	41	2	17
	AR	2	40	2	18
	AZ	1	41	3	17
	IA	2	40	2	18
	FL	2	33	2	25
	DE	0	53	4	5
80% Proficient	AK	2	67	4	20
	AR	2	68	4	19
	AZ	1	72	5	15
	IA	2	66	4	21
	FL	2	62	4	25
	DE	0	82	6	5

IN GENERAL:

- As AMO increases more schools benefit from both safe harbor and growth model.
- Very few schools meet the criteria for both safe harbor and growth model. (Yellow column)
- More schools are meeting the growth targets than meeting the safe harbor targets. (Green column is greater than blue column)



Study 2

- Compare accountability decisions for five models, applied to the same state data (different state than Study 1)School Accountability Models
 - Status model
 - Growth toward standard – Projection model
 - Value table (different than GMP)
 - Conditional growth percentile methods
 - Cumulative effects value added model



Study 2 – Quartile Correlations

Correlation Coefficients on School Quality Scores among 5 Accountability Models

Models	04 Status	05 Status	06 Status	NVT	Projection	CGP	CEM
04 Status	1	0.80	0.79	0.30	0.71	0.05	0.47
05 Status		1	0.85	0.20	0.89	0.08	0.63
06 Status			1	0.51	0.84	0.34	0.72
NVT				1	0.20	0.65	0.50
Projection					1	0.10	0.65
CGP						1	0.56
CEM							1

Study 2 – Decision agreement

Pairwise Kappa Statistics on Ranking of Schools among Different Models

Models	Status	NVT	Projection	CGP	CEM
Status	1	0.43	0.66	0.25	0.55
NVT		1	0.21	0.50	0.42
Projection			1	0.09	0.47
CGP				1	0.41
CEM					1

Study 2 – Overall decision agreement

Fleiss Kappa Statistics and Kendall's coefficient of concordance on Ranking of Schools Across All Models

Categories	Kappa	Standard Error	Z	Prob> Z
1: Fail	0.62	0.01	46.74	<.0001
2: Low Pass	0.21	0.01	16.37	<.0001
3: Pass	0.23	0.01	17.43	<.0001
4: High Pass	0.55	0.01	41.69	<.0001
Overall Fleiss kappa	0.40	0.01	40.73	<.0001

Kendall's coefficient of concordance (W) = 0.61



Why the overlap between Status and Growth in GMP states?

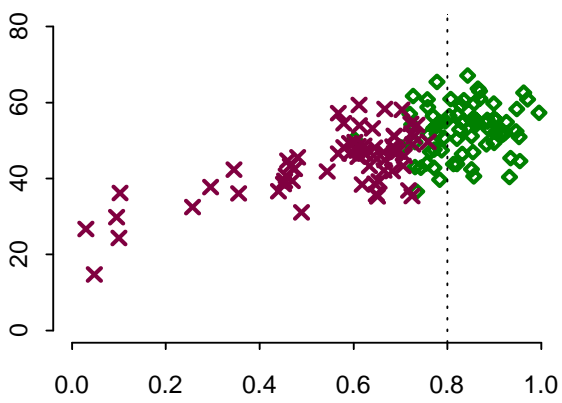
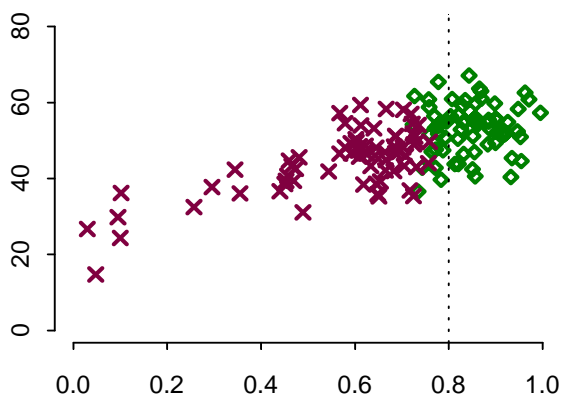
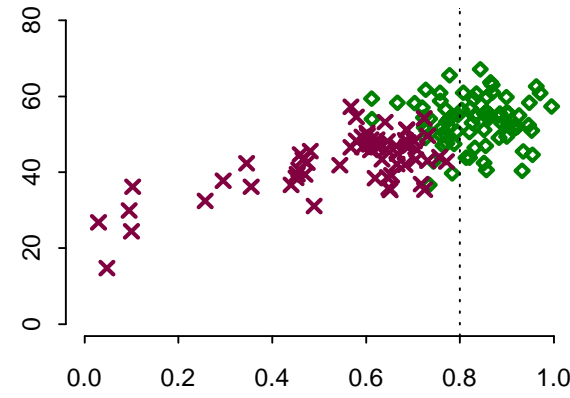
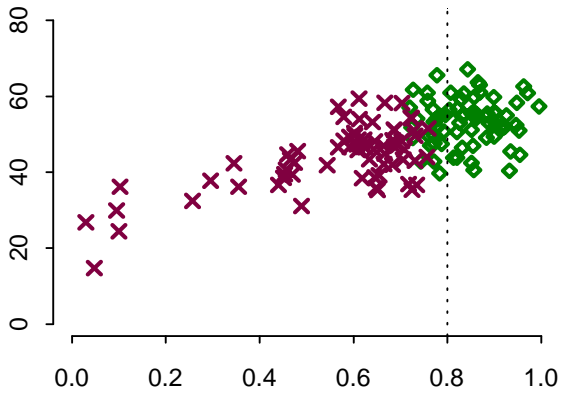
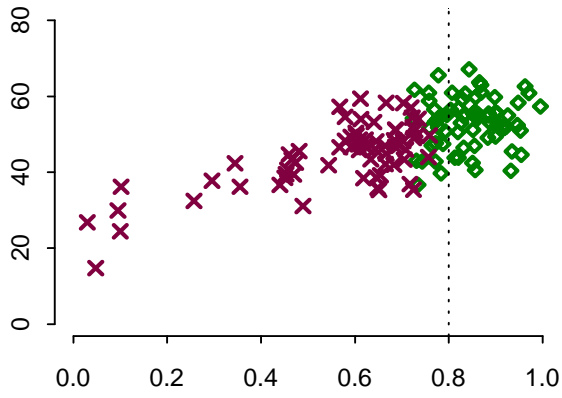
- Some approaches highly influenced by number of students already proficient
- GMP criterion of 3-years-to-proficient so steep that most schools do not meet it, especially schools that start low in Status



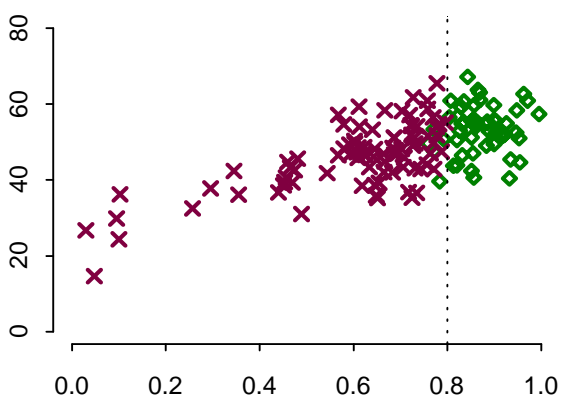
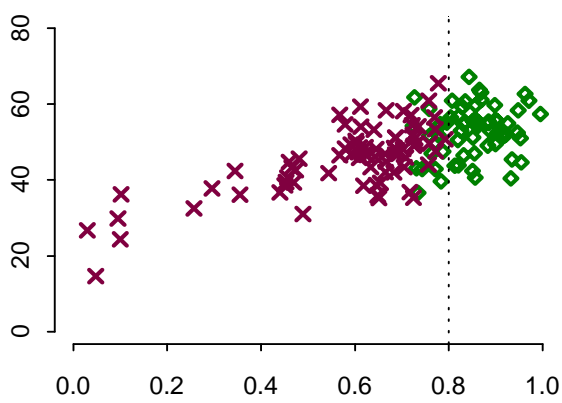
Why? – Some details

- Proficient-Plus approaches – where calculate Growth only for the non-proficient students – large percentages of students proficient overwhelm small additional number who make growth
- Proficient-Growth approaches – large numbers of proficient students make enough growth
- Growth-but-not-to-3-year-standard – steep trajectory means lower students/schools may increase scores but not enough – need to be close to Status AMO already to make AYP





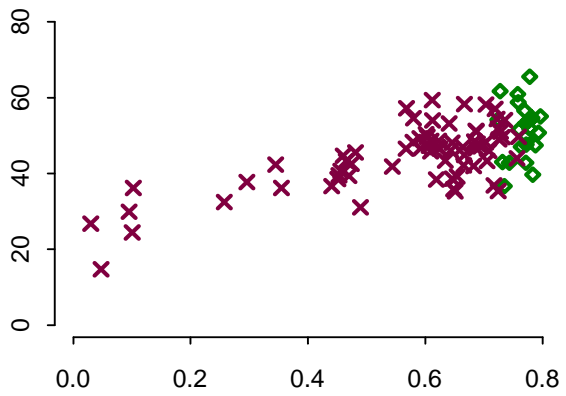
— Meets Growth Model
— Does not Meet Growth Model



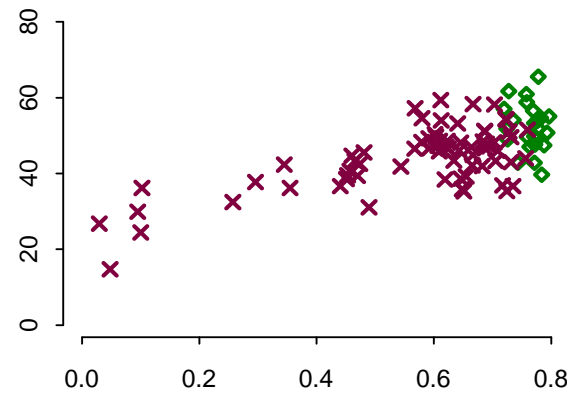
Percent Proficient

Average Growth Percentile

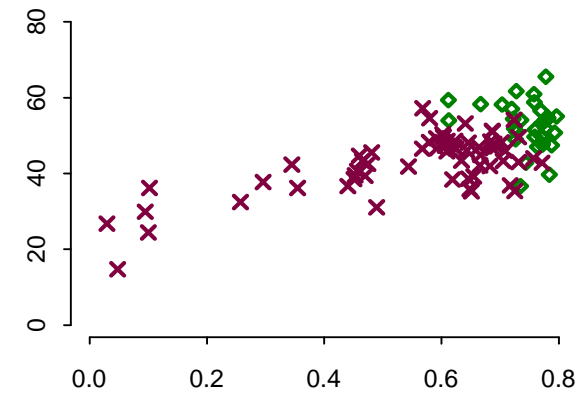
Average Growth Percentile



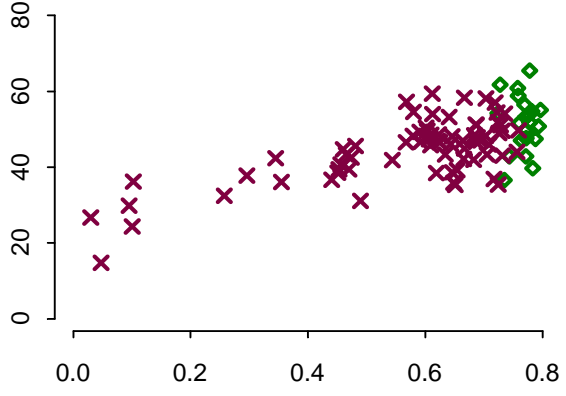
AK Growth Model



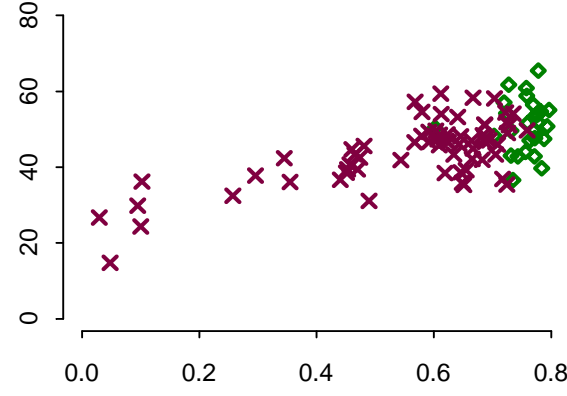
IA Growth Model



NC Growth Model

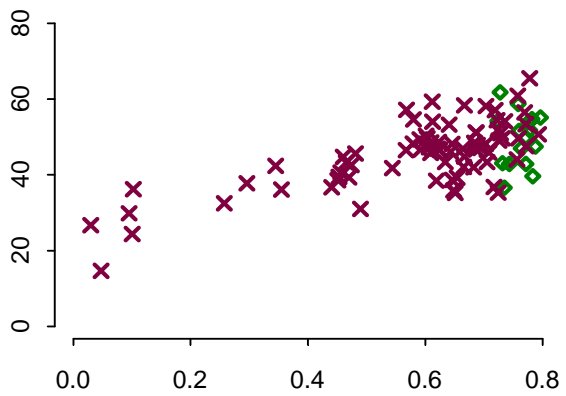


AR Growth Model

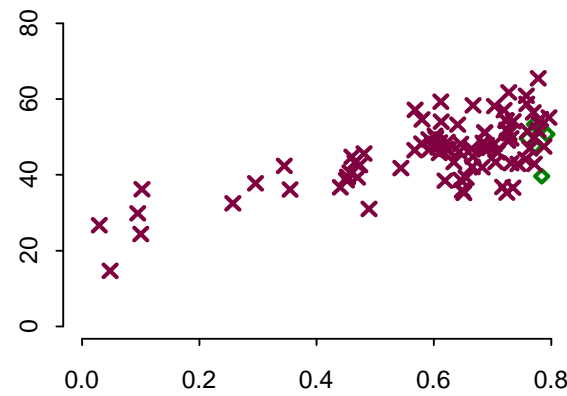


FL Growth Model

— Meets Growth Model
— Does not Meet Growth Model



AZ Growth Model



DE Growth Model

Status Score

But... look at interplay of all factors

- Consider Safe Harbor (e.g., Iowa's 2007 results)
- A very small number of students can trigger safe harbor in small schools
- Example:
 - School has 50 total students, 50% proficient in 2006
 - To meet Safe Harbor in 2007, need 10% reduction in non-proficient (5% increase in percent proficient, or 3 students)
 - With a subgroup of 30 students, 10% reduction = 5% of 30, about 2 students
 - How much of this is error/unreliability?



A sample transition table

Yr 2 →	1	2	3	4	5
1	66	19	15	0	0
2	25	31	42	1	0
3	4	11	70	13	3
4	0	0	34	40	26
5	0	0	7	24	69



Design Factors to consider

- Purpose
- Whom to measure
- How to measure
- How much is enough
- How to combine in other performance indicators to make accountability decision



Whom to measure

- All students
- Non-proficient students
- Non-proficient students in low status schools



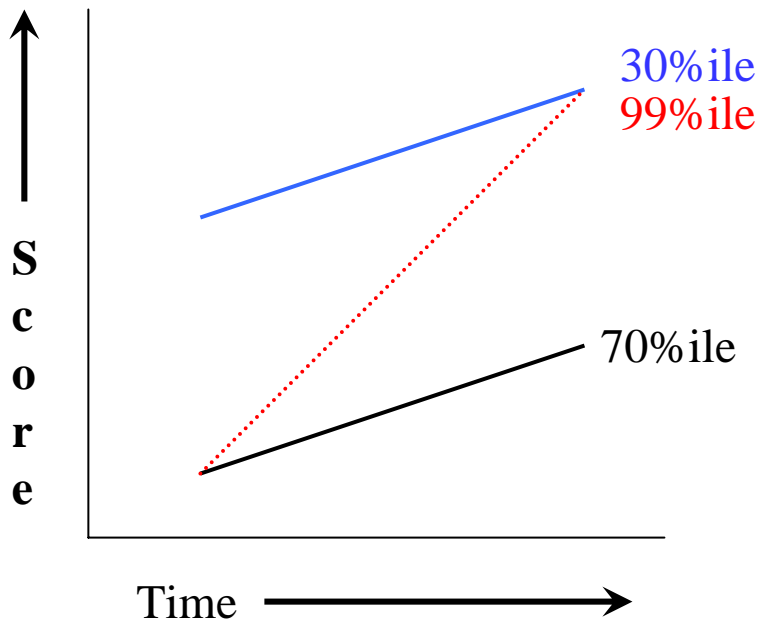
How to measure

- Growth model
 - Change scores
 - Value tables
 - Projection models
 - Traditional growth models
 - Growth percentiles

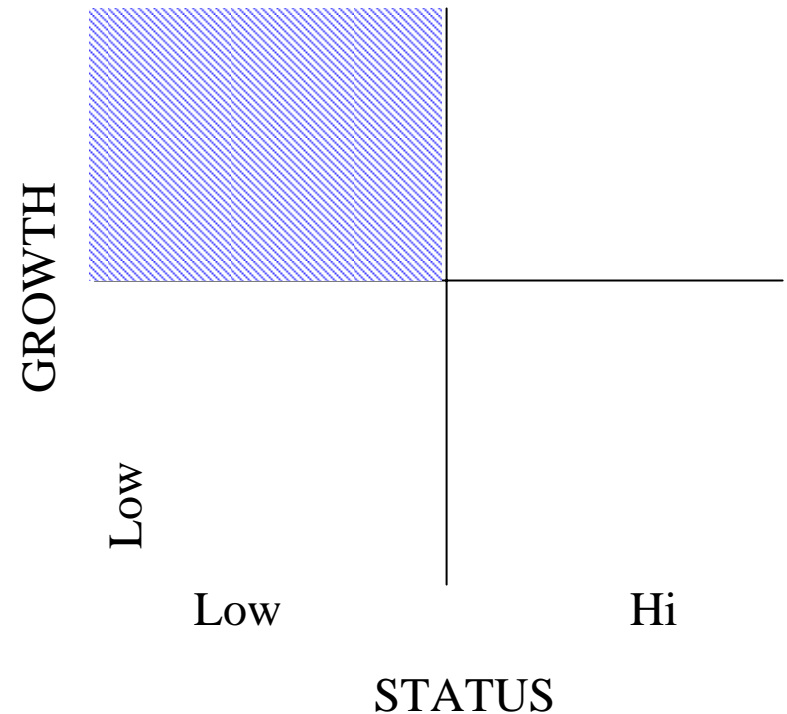


How much is enough?

“Effective school”
for a student



Status vs. Growth



How to combine into accountability

- Status only
- Growth only
- Cumulative Student Status and Growth
- School Status and/or Growth



For more information:

Center for Assessment

www.nciea.org



Jenn Dunn

jdunn@nciea.org

Brian Gong

bgong@nciea.org

