

Using Student Growth Measures in School Accountability

Brian Gong and Rich Hill
Center for Assessment

Presentation at the 2004 Reidy Interactive Lecture Series
Nashua, NH October 8, 2004



Recommendations

- Simple growth models most appropriate for school accountability
 - One view of several
 - Could be NCLB-compliant with some changes in USED interpretation and in statute
- Should define “expected growth” using policy informed by data
- More complex, “conditioned” value-added models less appropriate as main models for school accountability
 - very useful for program improvement
 - may be useful for supplemental accountability



Student Growth/Value-added Questions

- “It’s much more important to me to know how much each student is learning – how much they are improving – than ‘how high’ they are. Continuous improvement should be expected of every student.”
- “It only makes sense to hold a school accountable by tracking individual student progress, because the ‘good class, bad class’ effect of successive groups is so large it’s like comparing apples and oranges.”



Growth/VAM Questions – 2

- “Our school serves students who come in disadvantaged and behind. We help them learn a significant amount each year. We’d like credit for that, even if they don’t all reach ‘proficient’ that year.”
- “We serve a significantly disadvantaged population – poor, minority, mobile. It’s not fair to expect these kids to learn as much in a year as rich, suburban schools. We’d like to be compared with schools with similar challenges.”



Growth/VAM Questions – 3

- “Most of how much kids learn is out of the control of my school/me as a teacher. Innate ability and motivation, home influences, stuff from their previous teachers all determine how much I can help. Just hold me accountable for how much I contribute on top of that.”
- “We (the legislature) would like to be assured that for every additional dollar being put into education, we’re getting a fair return in learning.”



Growth/VAM Questions - 4

- “I’d like to be able to track my students’ progress as they learn during the year. I’d especially like to know if they didn’t know something so I could help them learn it better.”



Why Use Student Growth for School Accountability

- Another natural unit
 - Same student learning over time (“How much did student learn this year?”)
 - Complements other “natural units”: class, grades, schools, districts
- Attribution and program evaluation
 - Amount “under school’s/teacher’s control”
 - Teacher evaluation
 - Relative (comparable) performance
 - Output per input
 - On-going assessment



Key Presentation Topics

- Design purposes: Accountability
- Performance views: Status, Improvement, Student growth
- Setting expected student growth for accountability
- Accountability and analysis: multiple layers
- Student growth accountability and NCLB
- Implementing student growth in accountability
 - Explicitly valuing growth
 - Reliability and validity of growth approach



Design Purposes: Accountability

- Accountability: Designed to influence behavior
 - Reflects shared values: important indicators, outcomes, etc.
 - Embodies clear, attainable targets and goals, known before action
 - Provides useful feedback
 - Has meaningful incentives aligned with desired behaviors/outcomes
 - Actors (students, educators, state) have appropriate control (e.g., “I can influence/what I do matters,” “System will respond,” “Rules are fair”)
- Insufficient on its own to bring about reform



Design: Related purposes - 2

- Assessment – What?
 - “Was student proficient?” “How many students in school were proficient?” “How much did students improve?”
- Accountability – Enough? & So what?
 - “Schools will receive zero points for students who don’t participate.” “Were *enough* students in the school proficient?” “If not, what should happen?”
- Program Evaluation – Why? Who? (attribution)
 - “Do students in Class A learn more than students in Class B, all other things being equal?” “How much of learning was due to program/person X?”
- Research “truth” – How? & Invariance
 - “How did instructional program Y help students learn math concepts A, B, & C?” “How did student solve problem Z?” “How true would this be for all other students/teachers?”



Accountable for What?

- Three views of performance:
 - Status
 - Progress
 - Improvement (successive groups)
 - Student longitudinal growth



Focus of Three Views

- Status
 - How high do students score on state assessments? What percentage of students were proficient?
- Improvement (Successive groups)
 - Is the school improving, or increasing the performance of classes of students over time (e.g., grade 3 this year higher than grade 3 last year)? Is the percentage of students meeting the state standards increasing each year?
- Student growth
 - Are students learning as they progress through the grades? Are individual students making expected progress from grade to grade?



Calculating Status

Year	Grade 3	Grade 4	Grade 5	Grade 6
2001	Status => Count or Avg. across grades			
2002	Status			
2003	Status			
2004				



Calculating Improvement

Year	Grade 3	Grade 4	Grade 5	Grade 6
2001				
2002				
2003				
2004				



Calculating Student Growth

Year	Grade 3	Grade 4	Grade 5	Grade 6
2001				
2002				
2003				
2004				



Three Views – Data Needs

- Status
 - Annual assessment, representative group (not every grade)
- Improvement (Successive Groups)
 - Annual assessment; representative group, comparable across years; consistent performance standards; “headroom” or accountability system that allows for ceiling effect; two years’ data minimum
- Student Growth
 - Annual assessment, successive grades; means to track individuals across time and schools; consistent performance standards/interpretation of growth; assessment that is sensitive to growth; two years’ data minimum (at least three for more complex models); student background data (including teacher, school assignments) if conditioning is used



Three Views – Analysis Needs

- Status
 - Easy analysis; but challenging bookkeeping: Account for each student (by subgroup); special cases (e.g., FAY, 95% participation unless $n \leq 40$, subgroup minimum- n ; alternate assessment achievement levels)
- Improvement (Successive Groups)
 - More complex analysis but transparent (can do with four function calculator): index, school growth targets, etc.; bookkeeping magnified by multiple year issues (but no tracking of individual students)
- Student Growth
 - Ranges from simple to highly complex; requires special analyses to set up “baselines” for all but the simplest growth models; most require specialized software and personnel; may not be easily auditable; data needs may be much more extensive; analyses and reports complicated by dealing with missing data; substantially more time to process data (?) and produce reports



Accountability Influences Behavior

Since an accountability system should influence behavior constructively:

- Each model (Status, Improvement, Student Growth) should:
 - Allow students/educators to have appropriate control (e.g., “I can influence,” “System will respond,” “Rules are fair”)
 - Reflect shared values of important indicators, outcomes, etc.
 - Embody clear and attainable targets and goals, known before action
 - Provide useful feedback (reports, etc.)
 - Offer incentives aligned with desired behaviors and outcomes



Evaluating Student Growth

- Measure: Time 1, Time 2, (Time 3, etc.)
- Calculate Change (Time 1, Time 2)
- Compare to some “expected growth”
- Vertical scale
- Pseudo-vertical scales
- Vertically aligned content and performance standards



Two Sources to Inform “Expected Growth”

- Data-driven estimates of “historical” growth (what is or what has been)
- Policy-driven growth targets (what should be)



Data-driven Estimates of Growth

- Use measurement of past performance to estimate where student should perform in the present or future
 - Or to attribute growth between two points to certain variables
- May use more simple to more complex models
- All address future in terms of past performance
 - What HAS been
 - NOT necessarily what CAN or SHOULD be
- Should be sensitive to context and time
- Reflects current disparities in performance between groups (what is)



Examples of Data-driven estimates of growth

- National/state trend line over time
- Selected subpopulation trend line
- Regression line (statistical pattern smoothing)
- Regression line, conditioned on variable(s) (VAM)
- Norms (e.g., “one year’s ‘normal growth’ for reference group”)
 - Statistical corrections, e.g. for regression to the mean, sampling error
 - (Linear vs. non-linear treatments)



Drawbacks of Data-driven Expected Growth

- Metric for measuring growth often not related to achievement levels
- Usually will not get many students to “proficient” over time
- Will result in different expectations for different groups (reifies past performance differences)
- Growth metric often a “black box” instructionally



Policy-driven Growth Targets

- Anchored on a long-term goal defined as valuable by beliefs, sustained by social agreement (not inherent)
- Explicitly considered for significant performance units (e.g., subgroups)
- (See Doran, Linn, NWEA, Hill/Gong/DePascale)
- Could be non-linear, vary by content area (individual pupil plan)



Examples of Policy-driven student growth targets

“Proficient” by target time or grade (e.g., high school, gr. 8)

- Start from baseline
- Calculate gap, divide by time units
- Set “expected growth” per year
- Metric matters! (Technical, communication, instructional action)

Vertical scale:

- Start at 220 in grade 4,
- Goal is 460 in grade 10,
- Then need 240 scale score points total over 6 years,
- So expected growth amount is 40 points per year, and yearly growth targets are : 220 in grade 4; 260/grade 5; 300/grade 6; 340/grade 7; 380/grade 8; 420/grade 9; 460/grade 10



Policy-driven Expected Growth - 2

- Pseudo-vertical/transformed scales: z-score transformations by grade, centered on proficient
 - Grade-level proficient is 370, 470, 570, etc.
 - Start: 320
 - Goal: proficient/on-grade level by grade 8: 870
 - “Keeping pace = 100 points per year”
 - Gap: 50 points (plus yearly growth) over 5 years – 10 points per year
 - Yearly growth targets: 320 in grade 3, 430/grade 4, 540/grade 5, 650/grade 6, 760/grade 7, 870/grade 8
 - Don’t need single vertical scale; can mix tests; adjust for incomplete population



Policy-driven Expected Growth – 3

- Vertically articulated performance standards (achievement levels)
 - Start: grade 3 Below Basic
 - Goal: grade 5 Proficient
 - Expected Growth: two achievement levels
 - Yearly expected growth targets: Below Basic in grade 3, Basic in grade 4, Proficient in grade 5
 - Note: Could create sub-achievement levels (Basic+)



Drawbacks of Policy-driven Student Growth Targets

- Usually not reflective of general current practice; higher than empirical
- Feasibility often unknown
- May not be as technically rigorous, or not have (yet) well-known statistical properties
- Systems only now emerging
 - Current state student growth/VA systems generally data-driven (NC, TVAAS)



Policy-driven Growth Targets Informed by Data

- Expected growth should reflect
 - Clear, desirable long-term policy goal
 - Informed by data
 - What is possible
 - By whom
 - Under what conditions
 - » E.g., Linn, 75th/25th %ile, “beat the odds”
- Subject to monitoring and modification
- Can do incremental data-driven informed by policy vision



Expected Growth: “On Track” to Target

- Policy-driven growth target: student is “on track” to achieve the target (e.g., proficiency) within defined time
- Has to be extended for proficient+ students
- Different than Status and Successive Groups
 - Student may not be proficient until last (target) year
 - Students’ expected growth may be difficult to relate to standards and instruction (e.g., vertical scale scores)
 - Need individual growth target for each student
 - Student’s growth target may be recalculated annually
 - Issues: multiple time points, error, regression, non-linearity



Policy Positions: Growth distributions

- Exercise

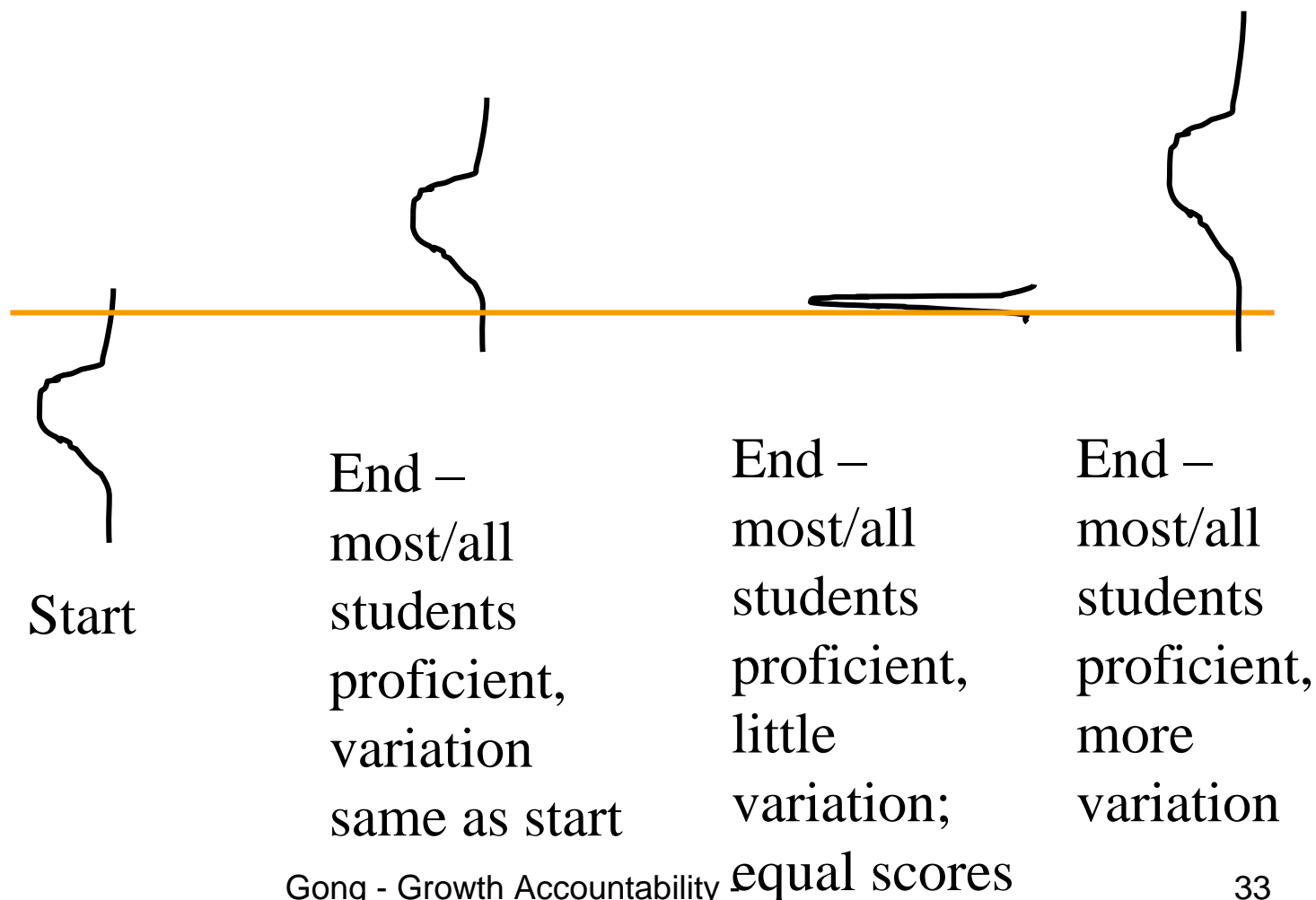


Expected Growth Targets for Schools

- Growth to goal or continuous (relative) improvement
- “Closing the gap”
 - “All student proficient” vs. “All students at least proficient”
- Relation to distribution of quality teaching within/across schools



What is desired distributions of student scores – for schools; students in classroom?



“Reasonableness”: Reflects perspective

- The reasonableness of an accountability system (or components, such as growth target) reflects the person’s role in the system
 - For example, “state” and “local” perspectives
 - State: Status report is sufficient; Local: want student growth
 - All agree that “all students should be accounted for,” but state, district, and school may not agree on who is accountable



Accountability Layers

- Most recent accountability in U.S. has focused on state holding schools/districts accountable
 - Most state constitutions; legal entities
- Have always had other layers/models
 - Teachers' grades for students; Principals' evaluations of teachers; school boards' evaluations of superintendents, etc.



Accountability and analysis?

- What is the right level for accountability (by whom, to whom)?
- What is the right level for analysis information?
- Proposal: Most value-added models are appropriate *analysis* to inform principals and teachers, but at inappropriately detailed for *school accountability*.



Levels of Analysis and Accountability

Level	Accountable (up)	Analysis (down)
State	To legislature	Are schools/districts in the state on track to meet proficiency goals
District	To state	Are schools on track
School	To district	Are grades on track
Grade/Dept.	To school	Are teachers on track
Teacher	To grade/dept.	Are students on track
Student	To teacher	



Student Growth and NCLB

- “On track” to proficient may be consistent with intent of NCLB, not consistent with statute about Status
- “Safe harbor” statute language does not prohibit student growth models, although would need a change in regulatory interpretation to allow it
- Expected student growth can be made to converge (somewhat) with Status goal, unlike current interpretation of “safe harbor”
- Conditional student growth almost certainly not consistent with intent of NCLB subgroup provisions (but helpful program evaluation)



Student Growth and NCLB – 2

- Could keep Status, Improvement, and Student Growth separate and provide multiple views of schools
- Could merge into overall rating
- Not strictly compensatory – need different types of assistance



Recommendations

- Simple growth models most appropriate for school accountability
 - One view of several
 - Could be NCLB-compliant with some changes in USED interpretation and in statute
- Should define “expected growth” using policy informed by data
- More complex, “conditioned” value-added models less appropriate as main models for school accountability
 - very useful for program improvement
 - may be useful for supplemental accountability



For more information:

Center for Assessment

www.nciea.org



Brian Gong

bgong@nciea.org

