Local Growth Models for Accountability

A Value-Added approach for Local Education Agencies to

- Set individual student growth targets
- Incorporate student achievement with teacher evaluations
- Document return on investment for programs
- Support leadership and board governance

ECRA group
BRINGING DECISIONS INTO FOCUS

May 2011

Contents

The Promise and Challenge of Accountability Reform

Assessment and accountability, the most prominent legacies of the "No Child Left Behind" era of education reform, are likely to be persistent and pervasive fixtures on the American educational landscape for many years to come. Recently, federal and state initiatives have shifted the focus of accountability models from achievement status, the measure of achievement at a single point in time – e.g. making Adequate Yearly Progress (AYP) on a state exam – to achievement growth, the progress or pace of individual students or groups of students over the course of their schooling. A significant benefit to analyzing change in performance of individual or cohorts of students over time is the ability to also assess the *impact* of critical components of their education, that is, the effect personnel and programming have on the students served by the district.

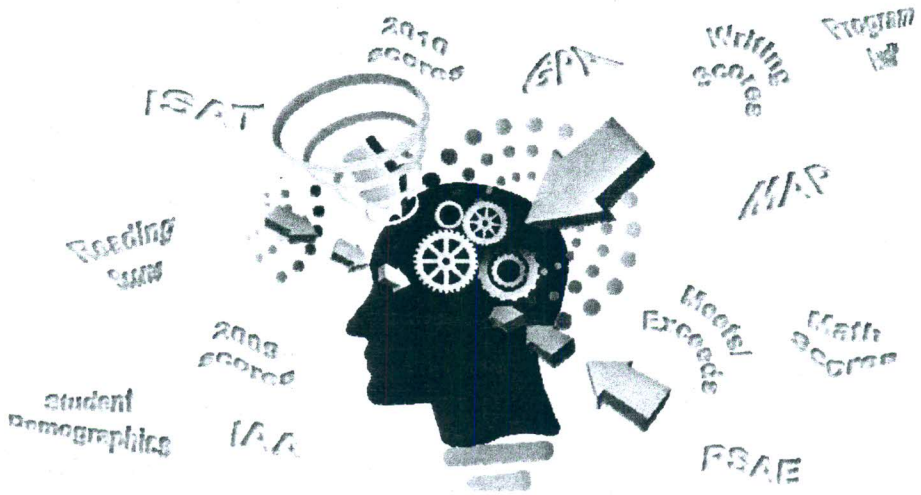
Growth models require a new analytic model and will allow school districts to re-align and enhance their instructional programs. Districts will need to employ a comprehensive, locally-driven analytic framework that focuses on current and future growth of a district's students in order to document individual student achievement, determine the effectiveness of instructional programs, assess the performance of school personnel and accommodate future reforms to education policy.

This white paper will describe the structure of this new analytic model, the process for implementing the model with local individual student assessment data, and the statistically sound ways in which this model informs decisions regarding students, personnel and programs. It will also provide district leadership with guidance to:

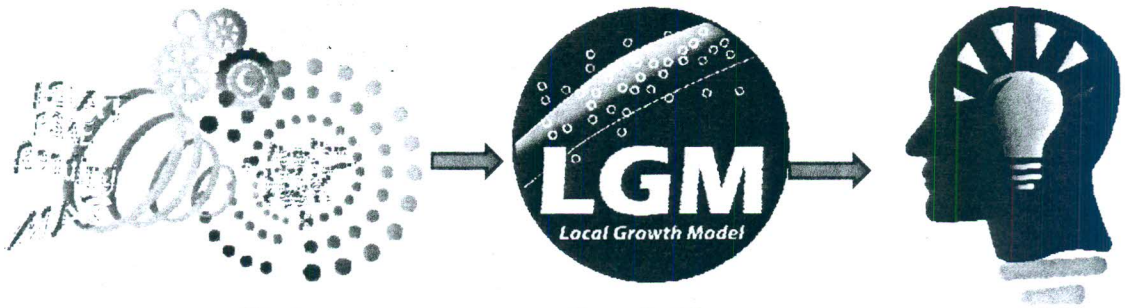
- **Transition from an achievement status model to an achievement growth model.**
- **Simplify the use of assessment data by distilling numerous data points to a single reliable measure of growth.**
- **Create control groups against which individual student growth can be benchmarked.**
- **Communicate achievement growth in assessment of student, teacher and program performance.**

The challenge in transitioning from an achievement status model to an achievement growth model is the analytic sophistication required to capture, integrate and document student growth accommodating the myriad of factors that affect student growth.

Old Data Paradigm: Interpret Unprocessed Data



New Information Paradigm: Draw inferences from processed data



Transitioning to a growth model requires that one adopt a new mental model for the relationship between data and information-based insights. Under the old data paradigm, educators had the burden of synthesizing and making meaning from a multitude of data elements, which can be unproductive, inaccurate and exhausting. Under the new information paradigm, assessment data is used to inform a model that distills multiple pieces of data into actionable information.

Growth Models and Value-Added Impact

Unlike "No Child Left Behind" (NCLB), which assesses whether a school achieves a minimum acceptable standard of performance, a *growth model* calculates progress or pace in order to summarize and project trends in student, classroom, and/or school achievement over time (Briggs & Weeks, 2009). By measuring the change in performance of individuals or groups of students in a given cohort, class or school, a growth model credits educational interventions, teachers and students for documented progress. Evaluation via longitudinal records of student achievement emphasizes individual student progress and provides schools with actionable information about which programs and teachers are "growing" students (Betebenner, 2009; Gong, Perie & Dunn, 2006).

Growth models address the following questions:

- What is a typical year's growth?
- How much growth actually occurred?
- How much growth should occur in order to reach a certain goal?

Effective growth models project what an individual student is likely to achieve at his/her typical growth rate and compares the projection to the student's actual achievement. In this sense, a well-designed growth model can be thought of as a mechanism for generating a pseudo-control group representing the probable achievement profile for a set of students. A well-designed growth model articulates the most probable student achievement profile. The comparison of probable growth with students' actual growth enables educators to evaluate the effectiveness of schools, teachers and programs.

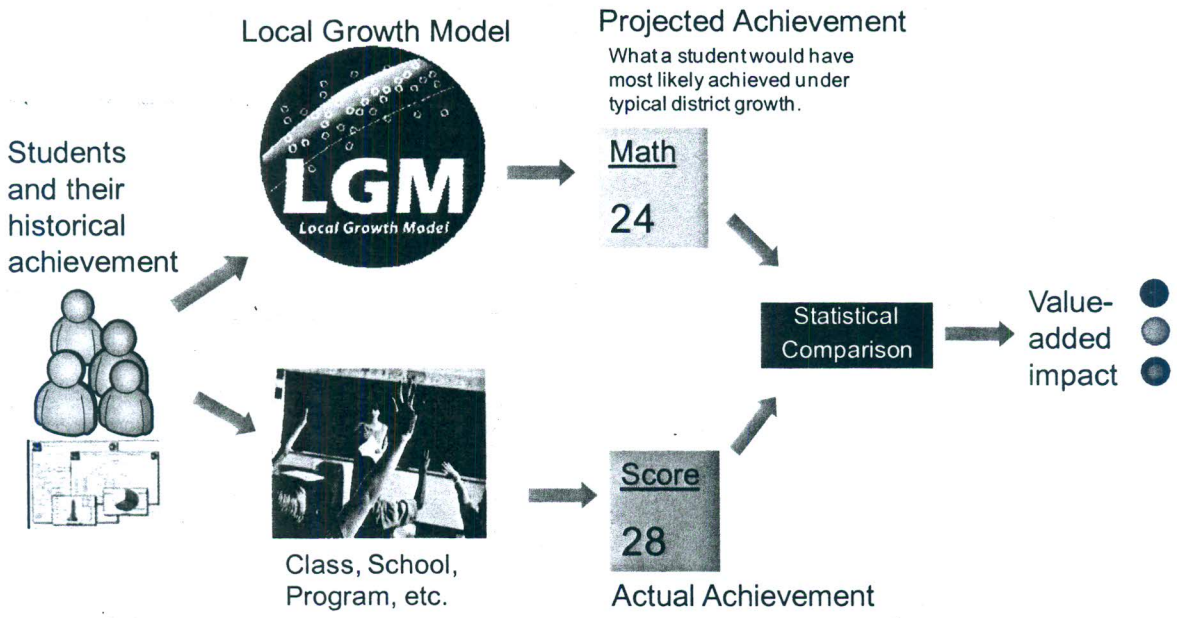
Growth models answer the following questions:

- Did an individual or group of students perform any better than they would have anyway?
- Which program or intervention is most responsible for the variance in student performance?

Historically, schools have been seriously limited in their ability to answer these questions because of the difficulty in establishing a control group against which to measure progress. By creating pseudo-control groups based on students' own longitudinal data, growth models allow districts to identify and quantify the impact that programs, interventions and teachers have on student achievement. Growth models provide data that links student achievement to teacher, administrator and program evaluation by revealing the value-added contribution to learning a teacher or program provides beyond what students would have demonstrated in the typical classroom, school or district.

The Logic of Growth Models

Growth models utilize historical individual student achievement data and analyze it using advanced statistical modeling techniques to predict a student, classroom, or school's expected future achievement under typical conditions. This projection is then compared to actual student achievement. The difference between projected and actual achievement is then expressed as the measure of growth.



While various statistical approaches exist for modeling the relationship between individual student's past performance and expected future performance, nearly all competing methods are built upon on a projected versus actual framework. Therefore, the choice of growth model depends in large part on the assumptions educators are willing to make regarding the relationships between curriculum and instruction, assessment and student performance.

Research on Growth Models

Growth models are only useful to the extent that typical achievement projections reflect a reliable and relevant benchmark for growth. Many states have highly localized school systems, each with unique demographic and performance indicators. Not surprisingly, there is no one-size-fits-all model that is effective for all districts: statewide accountability systems may or may not be a viable measure at the local level to identify what learning has occurred (Koedel & Betts, 2009). Still, there are several findings that researchers generally agree upon in regards to growth models.

Research indicates that:

Growth models are best suited for local measures and for initiatives nested at the district and school levels.

These provide the most actionable data of teacher, program, administrator and school effects (Choi, Goldschmidt & Yamashiro, 2006; Kelly & Monczunski, 2007; Linn, Baker & Betebenner, 2002; Raudenbusch, 2004). When growth is defined as what a set of students would have most likely achieved under typical district growth, the result is a locally driven framework built from district assessment measures, and is readily adaptable to multiple evaluation contexts.

The choice of growth model should balance statistical sophistication and transparency.

While statistical models must be sophisticated enough to capture reality with integrity, they must also be expressed in a manner that is relatively simple and informative to a broad audience (Amerein-Beardsley, 2008; Andrejko, 2004). Local growth models are purposefully developed to communicate educational growth and value added impacts through easy to read tables and graphs without sacrificing statistical complexity (Sanders, 2000; 2006).

Gain scores are unreliable.

Gain scores are the most commonly implemented method by which school systems determine the extent of the change due to an intervention or process. Gain scores simply take the difference between a pre-measure and a post-measure on a common scale. The unreliability of gain scores has been documented for decades, and continually emerges from the research community as a caution to educators. Despite being well documented, educators continue to rely on gain scores as measures of growth, perhaps as a result of the transparency and conceptual simplicity of gain scores. Nevertheless, the research community recommends that gain scores be avoided as the basis for growth models due to their instability. (Lord, 1956; Fortune & Hudson, 1984; Rachor & Cizek, 1996; Reckase, 2004; Sanders, 2006).

Most educational achievement scales are not true growth scales.

Research cautions against publisher claims of uniform growth scales. A more robust and flexible approach is to use analytic models that do not assume that assessments are on the same scale. Furthermore, gains on said scales are correlated with pretest scores. Therefore, expected growth on educational achievement scales depends largely on initial test scores, so an initially lower achieving student with a 10-point increase over the year has not learned the same amount as an initially higher achieving student with a 10-point increase. (Ballou, 2009; Reckase, 2004).

Growth projections should be based on at least 3 prior scores or time points.

In order to mitigate measurement errors and the influence of contextual or demographic factors, at least three prior scores from an individual student must be used to create the basis of expected future performance. Research has shown that when at least three prior scores on an individual student are used, the influence of other variables are largely mitigated to a negligible level. (Ballou, 2009; Sanders, 2000; Sanders & Wright, 2008).

Developing a Local Growth Model (LGM)

Designing and implementing a practical, psychometrically sound LGM can be simplified by reducing complex statistical models into a sequence of simpler approaches. The following sequence enables school districts to readily apply the model to new scenarios, programs, teachers and other accountability structures. Developing an LGM:

Step 1. *Identify the measures that will serve as the accountability system.*

Any growth model is only as good as the measures used to document achievement. Research literature makes clear that a strong accountability system requires a coherent set of learning standards, curriculum, pedagogical strategies and assessment programming that is well aligned to both the content and cognitive complexity of curricular objectives (National Research Council, 2010; Sanders, 2006; Sanders & Wright, 2008). So long as what is measured via assessments represents the scope and cognitive complexity of what is actually taught, school districts can use LGMs to determine relevant benchmarks for growth, project future achievement and quantify elements of the instructional program that provide value-added growth. Developing an effective growth model does not require a growth scale or the use of consistent metrics.

Step 2. *Use anchor years to serve as growth benchmarks and to develop a model that quantifies typical growth for an individual student.*

The LGM requires that typical growth profiles be documented locally. Decisions about anchor data are critical to the efficacy of the model. The research literature strongly supports the integration of multiple data points in order to account for students' prior achievement levels and their expected growth trajectories, as well as support inferences regarding school and teacher effects (Harris, 2008; Sanders, 2006; Sanders & Wright, 2008). To do this, a school district must use its own historical data to develop and calibrate the statistical models.

When choosing anchor years, it is important to identify data points that have the same achievement measures, implementation sequences and test administration windows that will be used in the future. The years must provide substantial data to calibrate statistical models.

Step 3. Use the model to project the most likely future achievement for each student.

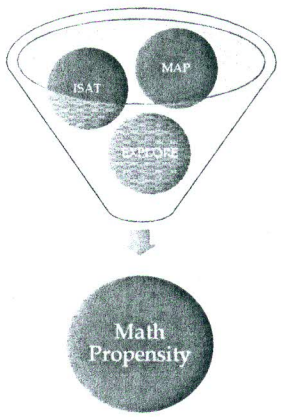
By integrating multiple data points reflecting individual student's achievement history and comparing these to scores obtained from other district students with the same test score history, a "Propensity Scale" for future performance emerges.

The *propensity score* is a composite of multiple prior individual student scores. It indicates the expected future achievement for a student based on his/her past performance. This provides a projected growth target that will serve as a benchmark to determine if actual growth is greater than, less than, or equal to the projection. In this manner, one reusable model meets the need for control group comparisons as well as predictability of actual performance.

Propensity Scores
Propensity scores reduce the complexity of assessment data by distilling all of the predictive information contained in multiple past test scores into a single score.

A propensity score captures the expected future performance of an individual student given his or her individual past achievement.

Multiple Math Scores Distilled to a Single Math Propensity



Step 4. Compare model projections to actual achievement.

This comparison yields the information on which student, teacher, program and district accountability rests. Having established valid, local projected growth trajectories for individual or groups of students, plotting actual achievement data will reveal the effectiveness of educational interventions.

The ability to attribute student growth to a specific teacher, program or intervention can be challenging due to the sheer number of contributing factors that may explain the academic performance of children. When attribution to one teacher is difficult due to the number of teachers assigned to a student, an optional, additional step is to employ a value-added model to report the individual teacher effect along with an adjusted effect for all teachers of a given student.

Applications of Local Growth Models

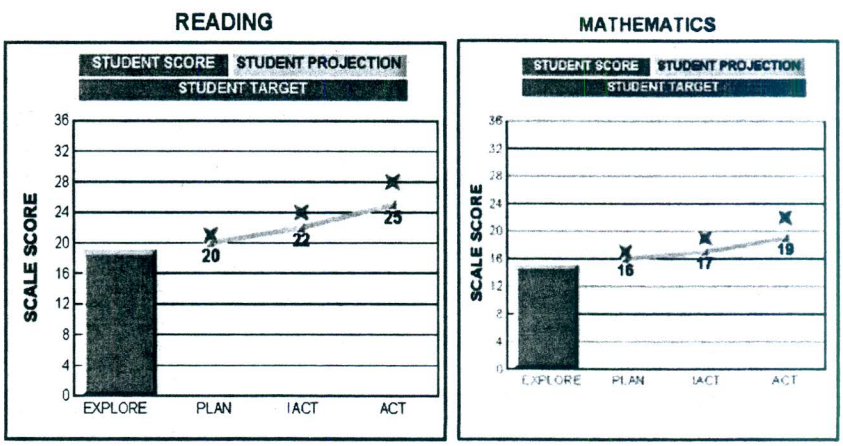
A local growth model can serve as a powerful tool for teachers, administrators and policy makers to initiate and sustain school improvement at all levels and to support student achievement. Information generated by the local growth model can be used to:

- Determine individual growth targets for students with or without IEP's
- Evaluate teacher, administrator, intervention or program performance
- Place students into appropriate course levels
- Project individual student academic performance
- Document return on investment for instructional programs
- Set cut scores for progress monitoring related to Response to Intervention (RTI)
- Support board governance

Set Individual Student Growth Targets

Local Growth Models enable districts to better analyze and monitor student achievement over time, and assess whether individual students are on track to reach important proficiency standards, e.g. grade level literacy or college readiness standards, in the future. This analysis can also result in modification to instruction and programming to address student needs. Individual monitoring and data collection can help determine appropriate course placement for students, monitor individual student learning plans, and determine long-term curricular goals and resource sharing across the K-12 continuum. It serves as a tool for identifying interventions and progress monitoring during the Response to Intervention (RtI) process.

The "Student Growth Reports" to the right illustrate how Local Growth Models can provide meaningful, relevant feedback to students, educators and parents by documenting a child's current performance as well as predicting future achievement. By providing personalized, individual 'road maps' for achieving academic goals, the "Student Growth Report" empowers students, parents, counselors and teachers to have shared accountability for student achievement.



Incorporating Student Achievement into Personnel Evaluation

Given the fact that current practice in teacher evaluation is based significantly on input variables such as instruction, assessment, classroom environment, content knowledge and professionalism and does not address the differences in teacher ability regarding output – i.e. student performance – it is time to revise our practice and pose the question:

"What impact is the teacher having on students and how do we measure it?"

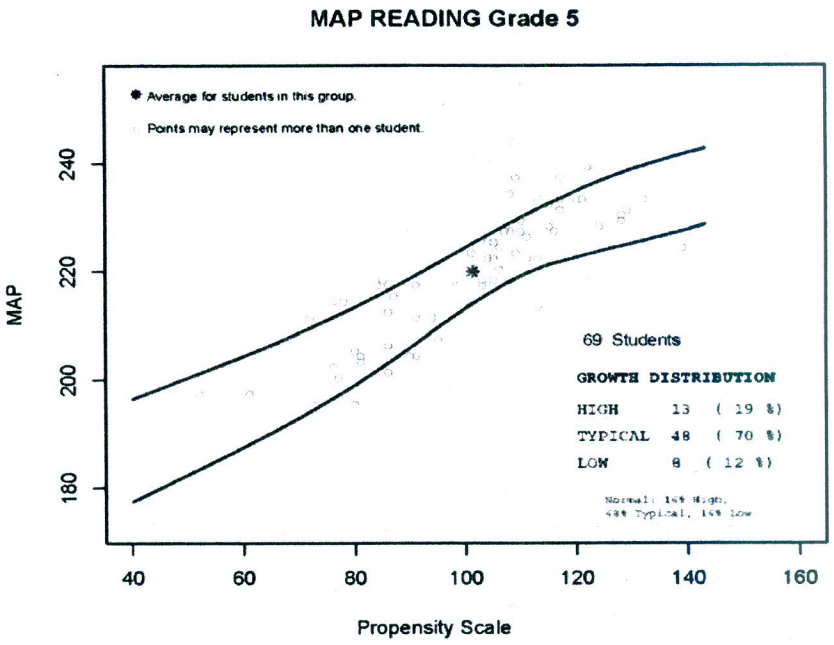
Local Growth Models offer a fair and transparent way to measure teacher impact on individual students. By accurately representing the growth trajectory for each student based on his/her actual performance history, teachers' performance is no longer dependent on how many of the 'good kids' are in their classes. LGMs can provide a valuable quantitative element to holistic assessment of faculty by calculating a "teacher effect" in terms of the difference between a teacher's predicted class average and the actual class average (Briggs & Wiley, 2008). The data generated can be applied in ways that improve instructional programming including:

- Identifying effective instructional and assessment strategies
- Determining needs in the area of professional development
- Celebrating effective teachers
- Identifying potential mentors and those who would benefit from coaching
- Replicating 'best practice' across content areas, grades and schools

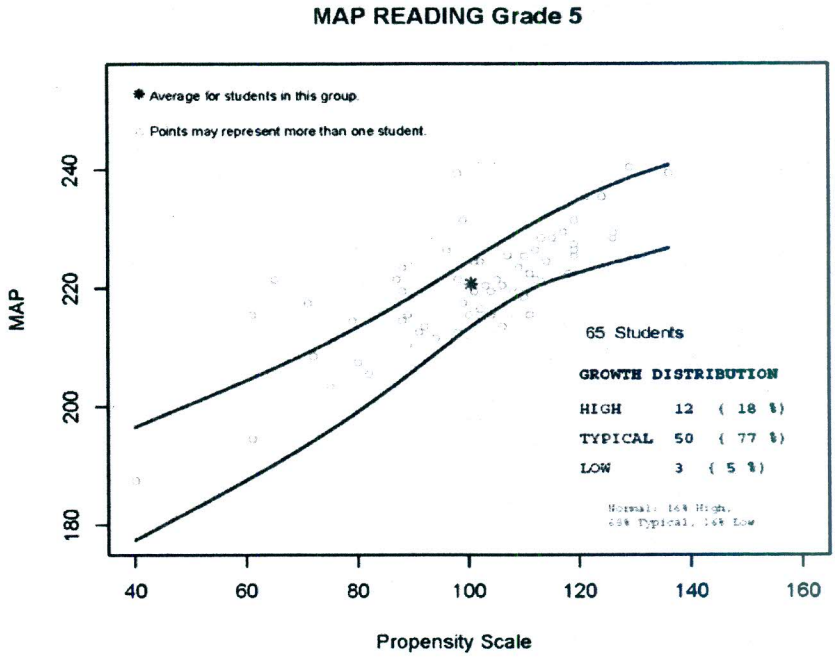
The following examples illustrate how Local Growth Models reveal the 'teacher effect' on student performance. By comparing the performance of students in other classes, the degree to which a teacher impacted student achievement - greater than, equal to or less than the average for the district – begins to emerge.

The graphs below depict how students' actual NWEA MAP test scores (Y axis) compare to the predicted propensity scale scores. The blue lines indicate one standard deviation above and below the projection. Under the expected growth, 16% of students should fall above the blue line and 16% should fall below the blue line. Each student is depicted by a circle.

The students in this fifth grade teacher's class show typical growth as compared to other fifth graders in the district (approximately 16% of students fall above and below the blue lines as expected). Based on the performance of the students, this teacher has demonstrated that he/she is effective in "growing" students consistent with predictions for average growth for grade five.



The students in this fifth grade class generally show growth above what is typical of fifth graders in this district, with only 5% of students falling below the blue line. However, the distribution is not necessarily uniform. The data indicate that this teacher is especially effective in 'growing' low performing students. Clearly, the conversations inspired by using a Local Growth Model will have a great impact on the teacher/ administrator evaluation process.



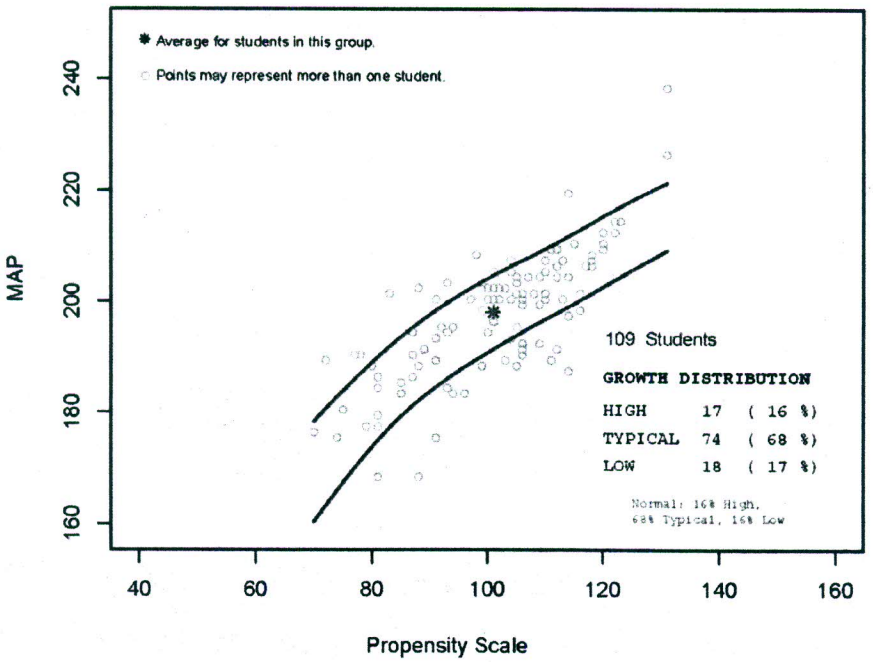
Monitor Return on Investment for Programs and Interventions

Evidence-based practice requires an accurate measure of performance to make reliable evaluations. Unfortunately, the growth and applicability of data-based decision making has been stymied by a lack of accurate information (OECD, 2008; Raudenbush, 2004). An LGM can quantify the impact programs and interventions have had on student learning, thus providing a framework to document return on investment. The incorporation of outcome variables into a district's evaluation model or strategic plan can provide greater insight into the degree to which long-term goals and school improvement plans have actually been accomplished.

The following two examples illustrate how an LGM can be used to determine the effectiveness of different instructional programs based on local student performance data.

The graph below shows the MAP Reading performance of students in the "21st Century Learning Program" as compared to the expected performance of the same students under typical conditions as predicted by the model. Approximately 16% of students fell above and below the two lines so it can be assumed that achievement growth under the new program was no better than under typical conditions and the district can now seek other curricular materials which may prove to be more effective in improving student achievement.

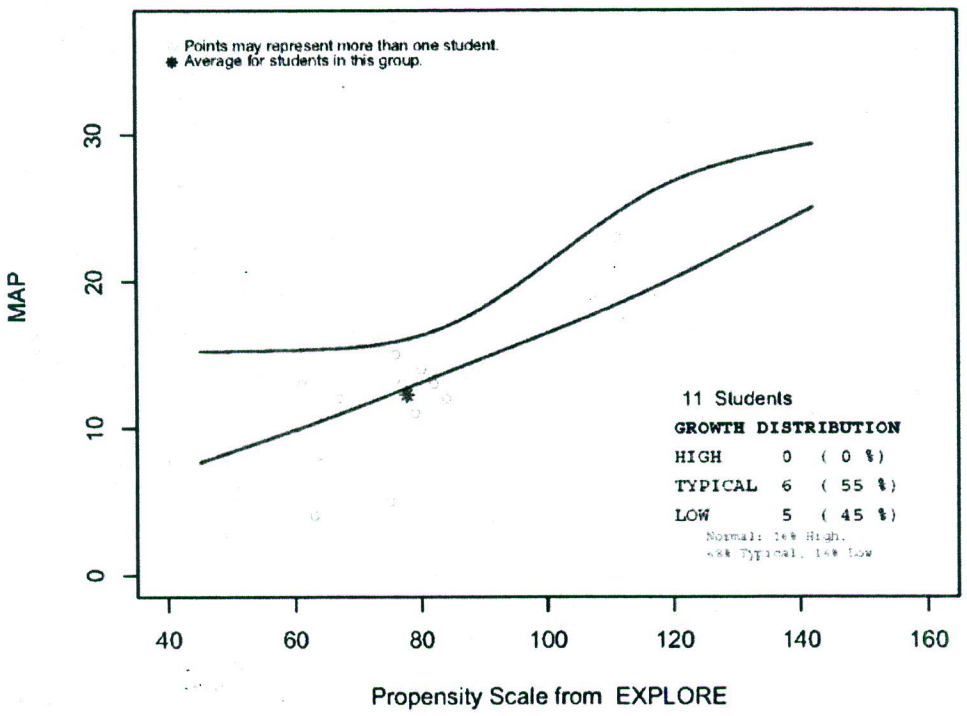
MAP READING Grade 2



On average, this group showed growth typical for students at this grade level in this district.

With the proliferation of interventions spawned by the demands of Response to Intervention (RtI), and the expense of maintaining a variety of grade appropriate and subject specific interventions, school districts more than ever need data to identify effective programs. The following example illustrates how the LGM identified a reading intervention pilot program that was not effective in "growing" students. The graph below depicts the performance of students in the "Pilot Reading Support Program" as compared to achievement predictions for the same kids had they not been part of the program.

MAP READING Grade 4



On average, this group showed growth below what is typical at the same grade level in your district.

By comparing students in the intervention to expected performance of the same students under typical growth, the performance of students with the intervention clearly shows that students in the "Pilot Reading Support Program" had lower than expected growth than they would have grown without the program. This is seen in the lower percentage of students above the top blue line than predicted by the model (0 percent vs. 16 percent expected) and the larger percentage below the bottom blue line than predicted (45 percent vs. 16 percent expected). As a result of the actionable data provided by the LGM, the district can make a decision to either end or alter the "Reading Support Program". Another ancillary insight in the graph is the presence of an outlying student on the top right of the graph who is in the program but may not actually need reading support based on his or her propensity score.

Support Leadership and Board Governance

Excellent school board performance must have a focus on measures that reflect strategic intent: *you are what you measure.* The school board and district management team should collaborate on which metrics they will monitor as indicators of a quality school system, and which metrics are important for the governance function of the board. Examples of these metrics are financial health, student achievement, program success, and class size, to name a few. A lack of alignment among conflicting metrics can result in an organizational battlefield for priorities and resource allocation. A forward-focused board, however, can narrow its focus to the levers of change and excellence by investing significant energy into identifying the few things that truly matter to measure – the structures and processes that actually "grow" students. Narrowing that measurement focus can effectively cut through information overload and free the board to create the capacity for performance excellence. LGMs can provide an accountability mechanism for school boards and district management teams to align and assess the district's mission through the actual practices of district personnel and students.

Example

Program	**# Students Served**	**Financial Allocation**	**Cost per Student**	**Achievement Return (value added: ACT points)**
Reading Support	112	$250,000	$2,232	0.8
21st Century Learning	248	$780,000	$3,145	0.1
Summer Academy	58	$112,000	$1,931	1.5

The example above illustrates how an LGM removes guessing, politics, or favoritism from discussions regarding resource allocation and provides clear, data-based information for effective board governance. By giving school boards the tools that reveal effective practices, effective programs, and effective teachers and administrators, an LGM empowers them to fulfill the board's mission of supporting the academic growth of all students through fiscally sound practices.

Discussion

The LGM is a reusable and scalable student performance analytic framework that provides reliable measures of program and personnel performance, learning targets for individual students, and actionable data to inform board decisions. Despite the escalating rate of change in the educational landscape, one aspect of the discussion remains constant. That is: the urgent need for a local analytic framework that can gather all the disparate pieces of data inherent in educational assessment, and use them to generate reliable, consistent, actionable information from which sound instructional and policy decisions can be made.

Your Local Growth Model turns your data into information and insights.

References

Amrein-Beardsley, A. (2008, March). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65–75.

Andrejko, L. (2004). Value-added assessment: A view from a practitioner. *Journal of Educational and Behavioral Statistics*, 29(1), 7–9.

Ballou, D. (2009). Test scaling and value-added measurement. Education finance and policy, 4, 351-383.

Betebenner, D. W. (2008). Toward a normative understanding of student growth. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 155–170). New York: Taylor & Francis.

Betebenner, D.W. (2009, December). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, 28(4), 42-51.

Briggs, D.C., & Weeks, J.P. (2009). The impact of vertical scaling decisions on growth interpretations. *Educational Measurement: Issues and Practice*, 28(4), p. 3-14.

Briggs, D. C., & Wiley, E. (2008) Causes and effects. In *The Future of Test-Based Educational Accountability*, K. Ryan & L. Shepard (Eds). Routledge.

Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). National board certification and teacher effectiveness: Evidence from a random assignment experiment (NBER Working Paper 14608). Cambridge, MA: National Bureau of Economic Research.

Choi, K., Goldschmidt, P., & Yamashiro, K. (2006). *Exploring models of school performance: From theory to practice*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007, December). Teacher credentials and student achievement: longitudinal analysis with student fixed effects. *Economics of Education Review*, Elsevier, vol. 26(6), pages 673-682.

Fortune, J.C., & Hutson, B.A. (1984). Selecting models for measuring change when true experimental conditions do not exist. *Journal of Educational Research*, 77(4), 197-206.

Fuhrman, S.H. (2010, April 7). Tying teacher evaluation to student achievement. *Education Week*. 29(28), Pages 32-33.

Gong, B., Perie, M., & Dunn, J. (2006). *Using student longitudinal growth measures for school accountability under No Child Left Behind: An update to inform design decisions*. Dover, NH: National Center for the Improvement of Educational Assessment.

Harris, D. N. (2008). The policy uses and "policy validity" of value-added and other teacher quality measures. In: Gitomer DH, editor. *Measurement issues and assessment for teacher quality*. Thousand Oaks (CA): Sage Publications; 2008. P.99-130.

Hershberg, T., & Robertson-Kraft, C. (2010, March). Rewards and supports. *The School Administrator. 3(67),* 28-31.

Hershberg, T., Simon, V. A. and Lea-Kruger, B. (2004, February). Measuring what matters. *American School Board Journal*.

Kelly, S, & Monczunski, L. (2007, June/July). Overcoming the volatility in school-level gain scores: A new approach to identifying value added with cross-sectional data. *Educational Researcher*, 36(5), 279-287.

Koedel, C., & Betts, J.R. (2007). Re-examining the role of teacher quality in the educational production function. Working Paper 07-08. Columbia, MO: University of Missouri.

Koretz, D. (2008, Fall). A measured approach. *American Educator*, 32(2), 18-39.

Linn, R. L, Baker, E. L., & Betebenner, D. (2002, August/September). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3-16.

Lord, F.M. (1956). The measurement of growth. *Educational and Psychological Measurement*, 16, 421-437.

McCaffrey, D.F., & Lockwood, J.R. (2008). Value-Added models: Analytic issues. *National Research Council and the National Academy of Education, Board on Testing and Accountability Workshop on Value-Added Modeling*: Washington D.C.

National Research Council. (2010). *Getting value out of value-added*. Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability. H. Braun, N. Chudowsky, and J. Koenig, *Editors*. Washington, D.C: The National Academies Press.

Organization for Economic Cooperation and Development. (2008). *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*. London: OCED Publishing.

Rachor, R.E. & Cizek, G.J. (1996). Reliability of raw gain, residual gain, and estimated true gain scores: A simulation study. *Paper presented at the Annual Meeting of the American Educational Research Association*, New York, NY.

Raudenbush, S. W. (2004, January). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics, 29*(1), 121–129.

Reckase, M. (2004). The real world is more complicated than we would like. *Journal of Educational and Behavioral Statistics*, 29(1), 117-120.

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005, March). Teachers, schools, and academic achievement. *Econometrica*, 73(2), p.417-458.

Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004, Sprig). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*(1), 103–116.

Sanders, W. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personal Evaluation in Education*, (14(4), 329-339.

Sanders, W. (2006, October). Comparisons among various educational assessment value-added models. *Paper presented at the National Conference on Value-Added*, Columbus, OH.

Sanders, W.L., & Wright, P.S. (2008, November). A response to Amrein-Beardsley's "Methodological Concerns about the Education Value-Added Assessment System." Brief. SAS: Cary, NC. Available at http://www.sas.com/govedu/edu/services/Sanders_Wright_response_to_Amrein-Beardsley_4_14_2008.pdf.

Yen, W. M. (2007). Vertical scaling and No Child Left Behind. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 273–283). New York: Springer.