2005

# Benchmarks and growth and success... Oh, my!

G Gage Kingsbury
*University of Minnesota*

# Benchmarks and growth and success … Oh, my!


## G. Gage Kingsbury

*University of Minnesota*


G. Gage Kingsbury (Ph.D., Psychology, University of Minnesota, 1984) is the Director of Research for the Northwest Evaluation Association (NWEA). He served as a member of the NWEA board of directors for seven years. His primary area of focus is in the application of Item Response Theory to practical assessment applications. Since developing his first computerized adaptive test in 1976, Gage has designed adaptive achievement tests that are currently in use by over 1000 agencies throughout the United States. This includes the development of the first adaptive test used operationally in K-12 education. In addition, he has developed procedures for adaptive testing that are currently in use in many operational adaptive tests used in selection, certification, and licensure, from military testing to the health professions.

Gage has published or presented over sixty studies dealing with item banking, item response theory, and computerized adaptive testing. He has served on the editorial boards for several peer-review journals dealing with measurement and assessment. Gage has also served as a developer of the American Council on Education standards for computerized adaptive testing and the Association of Test Publishers guidelines for computerized test development and use.


## Abstract

In order to inform decisions in our schools, information about student achievement has to be accurate and timely. The information also has to be presented in a fashion which encourages teachers and schools' personnel to make the best possible decisions. One of the most basic pieces of information concerns whether the school is doing a good job educating its students.

This paper will discuss some recent research concerning attempts in the United States to use student proficiency levels and content standards to identify schools that are struggling. It will also discuss a model that combines growth and standards to improve our ability to identify successful schools. Finally, it will discuss the use of an assessment system that fosters improvement in education.


As long as there have been schools, there has been the question of which school is the best. From sports teams to beautiful grounds to academic competitions, this question is discussed daily in coffee shops around the world. While it is clear that there is no 'correct' answer to this question, it is not for lack of trying.

In the United States, many folks think that public education is not doing as well as it might. However, these same folks will defend with all their might the quality of education and the quality of teachers at their child's school. The reason for this strong defence is simple. Parents can see how their son or daughter grows in school from day to day and from year to year. While they might not be able to quantify 'school success', they can see their daughter learning to read and growing into a person with profound capabilities and potential.

While the answer to the question of what makes a successful school is not an easy one, it is clear that it involves the amount that a school helps students grow in their knowledge, and in their love of learning. It seems clear that a model for school success that doesn't include the growth of an individual child is not a very useful model.

This paper will discuss some recent research concerning US attempts to use student proficiency standards to identify schools that are struggling. It will also discuss a model that combines growth and standards to improve our ability to identify successful schools. Finally, it will discuss the use of an assessment system that fosters improvement in education.

## Research on US attempts to identify struggling schools

The US federal government has used several approaches to identifying 'schools at risk' in the past. To use less loaded language, let's call this the 'search for schools that aren't very successful'. The current approach that the 'feds' are using to identify less successful schools is seen in the AYP (Adequate Yearly Progress) provisions of the No Child Left Behind Act. Under this legislation, schools are judged to be successful or not depending on the percentage of students in each grade and subgroup who can successfully reach a defined level of proficiency in reading and mathematics. The details of the level of proficiency and the content being assessed are left to the states to decide.

The approach taken in No Child Left Behind (NCLB) does not include the growth of individual students. Instead, it looks at the percentage of students who happen to be able to clear a single proficiency hurdle on a single test on a single day of the school year. While this

can be an important piece of information, it isn't the most important element to look at when measuring school success. Researchers investigating this issue have raised the following four concerns:

1 *Single point-in-time analyses may reflect demographics rather than effectiveness.* They cannot distinguish between schools that accelerate skills and those that allow students to languish. Cross-sectional measures do not tell us whether students entered with high or low skills or whether they have gained or lost ground as a result of instruction. Flicek and Wong (2003) characterise the cross-sectional percent-proficient model as one of the least valid evaluation methods. Schools that serve primarily English-speaking students who are not in poverty tend to have higher results. The data do not show which schools have been effective with the population that they serve (Kim & Sunderman, 2004b; Baker & Linn, 2002; Buchanan, 2004).

2 *The NCLB model does not take the performance of students above or far below the standard into account.* When the goal is to get the greatest number of students to meet the standard in a year, schools quite sensibly direct efforts at those performing just below the cut-off point. Schools earn no credit for improving skills of the lowest performing students or for getting gifted student to work to their capacity. Critics have pointed to this feature of NCLB as a disincentive to excellence, encouraging states to set low standards in order to concentrate on fewer students and look better in public reports (Marion et al., 2002).

3 *The current system does not necessarily lead to better placement for students in low performing schools.* The examples shown above indicate that students who move to schools with higher percentages of students meeting the standard may not get a better education. As Kim and Sunderman (2004a) note, students who take advantage of transfer opportunities afforded under NCLB often move from schools with support for low performing students to more affluent schools that do not have remedial reading programs, tutors or supplemental Title I money.

4 *Expectations of AYP need to be tempered by looking at observed results in exemplary schools.* In his 2003 address, as president of the American Educational Research Association, Robert Linn illustrated the gulf between NCLB expectations and observed performance. Using state and NAEP data from across the country, Linn projected that reaching 100% proficiency in twelve years would be highly unlikely. He called for the use of research to establish goals that are stringent, but feasible.

One of the primary outcomes of NCLB has been renewed discussion about what constitutes school success and what school accountability models should look like. Although the law and its implementation have not been straightforward nor without controversy, this extended dialogue and the associated research will definitely improve our knowledge of how schools work for students. Consider a person comparing the Adequate Yearly Progress of two schools and asking the following question:

> If two schools finish the instructional year with the same percentage of students above the proficiency levels established by my state department of education, are both schools equally effective?

A prudent person would probably answer this question 'I don't know'. We can't judge student growth by looking at a student's current level, and without knowing anything about student growth in a school, we can hardly judge whether that school is successfully educating its students. It is possible that some of the students in one school exceeded the state performance standards before they came to this school. Status relative to the performance standards is not sufficient to identify individual or school success. Both student status and student growth are needed to paint a complete picture of a school's effectiveness.

The graph below shows how students' fifth grade mathematics status (Average Score) and growth (Growth Index) compare in a group of several hundred elementary schools from throughout the United States (McCall, Kingsbury, & Olson, 2004). Several findings are clear from the graph, but the most important are the following:

- Schools with very similar status levels may differ greatly in the amount of growth they cause in their students (schools A and G, for example)
- Schools with cause vary similar growth for students with very different status levels (schools A and D, for example)
- A high-performing school may not be one where you would want your children enrolled (consider school F, for instance).

These findings mean that some schools are consistently more effective in causing growth for their students, regardless of the students they work with. This is important information about the success that a school is having with its students.

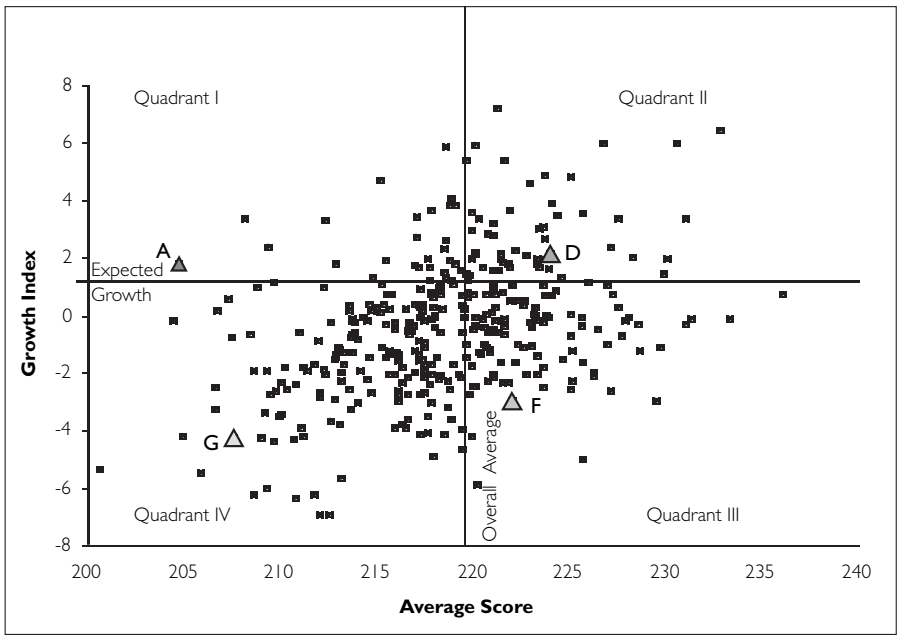


**Figure 1** Comparison of average mathematics scores and growth index for grade 5 students by school

It is clear that implementation of NCLB provides US schools with a variety of challenges, and many opportunities to make education better. Students and educators deserve to know what is expected of them, and states' efforts to set content standards and standards of performance have clearly helped schools bring greater focus to improving achievement. Pursuit of improvement requires that public policy, resources, and sanctions to be applied in a purposeful and prudent fashion.

This study makes clear that a key element that is not represented in NCLB metrics is **individual growth**. A more complete accountability system would reward schools for the growth they nurture in students. Proficiency standards are useful in measuring status, but they can create inequity by focusing schools on the relatively small number of students who are nearly proficient, and diverting their attention from those who are far from proficient.

## The Hybrid Success Model

An example of the category of models that include both growth and proficiency is Kingsbury and Houser's (1997) Hybrid Success Model. To measure success of a school with this model, we measure academic growth of each student in the school. To the extent that students are growing as much or more than expected and growing towards or beyond proficiency, the school can be judged a success. To determine this:

- Each student is given a growth target each year, in each content area of interest;
- The growth target, if achieved, will require every student to grow as much as a pre-defined comparison group;
- If the student is below the proficiency level, the growth target will be higher, requiring growth that will result in proficiency within a pre-defined period of time;
- Each student is assessed at least twice yearly, and the student's growth is calculated and compared to the growth target;
- The school gets credit toward success for each student reaching or exceeding their growth target; and
- The school is judged a success if its total credit exceeds a pre-defined performance level.

That is the entire process. It can be implemented in any setting that has defined curriculum standards and proficiency levels, and uses a measurement instrument that is vertically scaled. It allows every student to 'count' in the measurement of school success, by requiring that very high and very low achieving students continue to grow, and it leads every student to proficiency and beyond. While current legislation tries to help those students who are struggling, the HSM process judges school success by looking at the success of every student in the school. The use of HSM should create a climate with rigorous but attainable standards, to the benefit of all students.

## An assessment system that serves students

A high-quality assessment system must meet accountability requirements, but it also must serve the needs of each student enrolled in the schools. In order to achieve this goal, the system might include the following components:

- Content standards that are fairly complete, and flexible to change;
- Performance standards that can be measured along a stable scale that measures growth across grades;
- Performance standards that have

consistent meaning across grades and across subject areas;

- Accurate measurement of student achievement and growth;
- Reporting of results to teachers and administrators in a timely fashion;
- Measurement of student achievement that allows the identification of areas of strength and areas of concern;
- A procedure for changing instruction based on areas of concern and areas of strength;
- Measurement of school success that allows the identification of areas of concern and areas of strength;
- A procedure for using information about school success to change policy based on areas of concern and areas of strength;
- A model for systemic effectiveness that allows a school district to measure its improvement across schools; and
- A procedure to improve a school system based on information about systemic effectiveness.

A simple set of tools can be used to make the assessment system described above a reality. These tools enable an organisation to craft a strong assessment system. The system will be able to meet accountability needs and provide accurate information to students and teachers. The set of tools includes:

- *A measurement system that includes a stable, cross-grade measurement scale.* An example is found in the NWEA RIT scale, which has demonstrated stability over more than 20 years, and which allows detailed characterisation of a student's achievement against a map of skills that common to a wide variety of curricula.
- *Assessments that are targeted at each student's instructional level, not the middle of a grade range.* Targeted tests or adaptive tests provide the most accurate measurement available today.
- *A model for examining school success that incorporates both status and growth.* One such model that is currently in use is the Hybrid Success Model. It incorporates reasonable growth for each student as one aspect of success, and incorporates additional growth that will bring every student to the proficiency level as another aspect.
- *A reporting system that fosters the use of data to improve education.* A variety of models for systemic, data-based change exist, but each one depends on providing meaningful reports to the people who need them before they get stale.

## References


Baker, E. L. & Linn, R. L. (2002). *Validity issues for accountability systems. CSE Technical Report 585, 2002.* National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles: UCLA.

Buchanan, B. (2004). Defining 'adequate yearly progress'. American School Board Journal, February, 2004.

Kim, J. & Sunderman, G. L. (2004a). Does NCLB provide good choices for students in low-performing schools? Cambridge, MA: The Civil Rights Project at Harvard University.

Kim, J. & Sunderman, G. L. (2004b). Large mandates and limited resources: State response to the No Child Left Behind Act and implications for accountability. Cambridge, MA: The Civil Rights Project at Harvard University.

Kingsbury, G. G. & Houser, R. (1997). Using data from a level testing system to change a school district. In *The Rasch tiger ten years later:* Using IRT techniques to measure achievement in schools. Chicago, IL: National Association of Test Directors.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. Educational Researcher, Vol. 32, No. 7, pp. 3–13.

McCall, M. S., Kingsbury, G. G., & Olson, A. (2004). Individual Growth and School Success. Portland, OR: Northwest Evaluation Association.

Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S., & Sheinker, J. (2002). Making valid and reliable decisions in determining adequate yearly progress ASR-CAS Joint Study Group on Adequate Yearly Progress, Council of Chief State School Officers: Washington, D.C.