# Validity study of Pioneer RESA's *Insight* growth model
# FINAL Report May 2011

# Eric W. Crane


## SUMMARY


WestEd conducted a validity study of the *Insight* growth model developed by the Pioneer Regional Educational Service Agency (Pioneer RESA). By design, this study was a small-scale review of evidence of validity. The study looked at the model's validity in three contexts: for school improvement, for teacher assignment and evaluation, and for school and system accountability. The study found evidence that the model can suggest actions that are consistent with other school improvement decisions and can generate information that can be acted upon for school improvement, such as pointing out needs for professional development. The study found that the teacher-level measures produced by the model have validity on their face and confirmed the view of Pioneer RESA that in teacher evaluation, other evidence should always be used along with *Insight* reports. The study found evidence that the model may be able to inform teacher assignment decisions. Finally, the study found evidence that the model can rank-order schools and systems for purposes of accountability, including decisions related to rewards and interventions. If the model is to be applied for high-stakes purposes, however, additional validation work—including correlations to other measures and formal review of the model with stakeholders—is recommended.


## BACKGROUND

### Introduction and study background

In March 2011, Houghton Mifflin Harcourt contracted with WestEd to conduct a validity study of the growth model used by Pioneer RESA in the agency's *Insight* Data Framework. Pioneer RESA is located in northeast Georgia. WestEd agreed to explore the central question of the review, *Is the growth model valid for the purpose(s) intended?*, via interviews with Pioneer RESA staff, document review, and examination of *Insight* reports. This report documents features of the model, information about how results are used, other context relevant for examining validity, and WestEd's findings.

## Study methods and data sources

Consistent with the scope and duration of the study, three types of data sources were used. First, extended discussion with staff of Pioneer RESA provided a great deal of background information. Jeremy Williams, Evaluation & Assessment Director, was the study's main point of contact and source for a wide range of information about the model, its history, and its implementation in the schools and systems served by Pioneer RESA. Michael Catledge of the Evaluation & Assessment team provided additional background on the model and the reports that are generated by the *Insight* Data Framework. Dr. Sandy Addis, Executive Director of Pioneer RESA, provided additional information on the agency's interest in the validity study, including the agency's responsibilities to the field as the application of the model continues to grow.

In addition to staff interviews about the *Insight* growth model, the study looked at written material about the model. The written material reviewed for this study included a description of *Insight* on the agency's web site, a white paper about the model, and a PowerPoint presentation explaining reports.

The third type of data source examined was the actual reporting associated with *Insight*. To understand the reports more thoroughly, WestEd received a login and access code to the *Insight* system. In addition, Pioneer RESA staff provided a comprehensive file of statistical output showing teacher-, school-, and system-level change results for fifth grade mathematics.

Certain relevant data sources were not part of the study. Notably, the study did not conduct interviews with teachers, principals, system superintendents, or stakeholders other than service agency staff. In the context of this report therefore, any references to the opinions or views of teachers and principals are as reported by Pioneer RESA staff. In addition, while a range of growth reports were part of the study, there may be other relevant reports, worksheets, and other explanatory materials used in discussions of student, classroom, school, or system performance that were not reviewed. Finally, the study did not correlate results from the *Insight* growth model with other system-, school-, teacher-, or student-level data. (In some cases, correlational work conducted by Pioneer RESA is referenced.)

## *Insight* Data Framework and the model

[*Much of the material in this section comes from the publicly available white paper,* Insight: a growth model for Georgia.]

The growth model has been used continuously since 2005 in 13 northeast Georgia school systems.[1] Like other growth models, the *Insight* model recognizes *improvement* in—rather than

[1] Banks County, Dawson County, Franklin County, Gainesville City, Habersham County, Hall County, Hart County, Lumpkin County, Rabun County, Stephens County, Towns County, Union County, and White County.

*status* of—student performance. The absence of a vertically aligned test for Georgia's schoolchildren was a strong motivator of the *Insight* model. According to Pioneer RESA staff, in the years immediately following the authorization of the No Child Left Behind Act (NCLB), when unprecedented attention was focused on measures of student performance, many teachers had difficulty interpreting students' scale scores on the Georgia Criterion-Referenced Competency Tests (CRCTs). The scale score for a student may have declined from one year to the next, for example, but the student had progressed. Educators likewise found normal curve equivalents, or NCEs, to be difficult to interpret. Pioneer RESA staff reported that the school systems in effect applied a double standard to the interpretation of cohort data on student performance: when scores went up, this success was deemed to be due to the teacher; when scores dropped, this problem was due to a difficult cohort.

One metric readily understood by teachers was percent correct on the CRCTs. However, looking only at percent correct can be deceiving, since changes in percent correct may reflect changes in the difficulty of the test as much or more than student growth. The *Insight* model seeks to control for the difficulty of the test by looking at changes in each student's percent correct relative to a larger population, such as all students in the region or state. The model answers the question, *how does each individual student compare to the full cohort, twice?* In the reports examined for this study, the larger population was the set of students in the 13 school systems served by Pioneer RESA. These systems include 84 schools and roughly 5000 students per grade level. Since each student's percent correct is compared to the overall larger population's percent correct, the average student score each year under the *Insight* model is zero. Furthermore, since only those students who have a score in successive years are included in the larger population, the average of all students' difference scores is also zero. For more information about the model, please see the white paper (Pioneer RESA, 2011).

The data reports allow ready comparison at many levels within a system. Through an easy-to-navigate interface, student performance can be compared across schools, subjects, class periods, grade levels, teachers, and domains within subject. Comparisons can be further modified by applying filters to the students included in the analysis: it is possible to filter for specific values of gender, ethnicity, and student populations (students with disabilities, designation as gifted, students receiving free- or reduced price lunch, English learners, and receiving migrant education funding).

## VALIDITY

At its core, the examination of validity is an exercise in correlation: if applying a new model (or process or measure) leads to decisions and actions that correlate highly with decisions and actions based on known and trusted models, processes, and measures that support a purpose, then the new model is said to be valid for that purpose. Earlier, it was noted that the study did not

conduct any empirical correlation work; however, in this section, the decisions and actions generated by the *Insight* growth model are examined in light of known and trusted practices, to explore the model's validity.

**Intended purposes**

Understanding a measure's intended purpose is necessary for a proper discussion of the validity of the measure. Consistently in the interviews and document review conducted for this study, the *Insight* model's present and historical purpose has been to inform and stimulate school improvement discussions. In the past few years, according to Pioneer RESA staff, there has been some expanded application of the model to teacher assignment and evaluation. Furthermore, at present, there are scattered pockets of pressure across the systems served by Pioneer RESA to expand the application of the model to school and system accountability. Each purpose is a separate context for validity and merits its own discussion.

*Validity for school improvement discussions*

Discussions about student performance data can be powerful drivers of school improvement (Jamentz, 2001). Student performance overall on the state test is a highly visible accountability outcome, but overall performance lacks the detail necessary for school improvement. By contrast, the *Insight* growth model reports have the potential to highlight trends that would be difficult to detect without the tool. The Pioneer RESA staff, in their white paper, present data that show that use of the *Insight* model is correlated with school systems having decreased numbers of schools in the *Needs Improvement* accountability status, greater likelihood of having all schools meet adequate yearly progress (AYP) requirements, and higher graduation rates (Pioneer RESA, 2011).

In interviews, the Pioneer RESA staff described a sequential data analysis process that supports school improvement. In this process, the examination of high-level (e.g., system) results leads to questions that can be explored with a deeper examination of the data (e.g., across schools), and that this drilling down continues (e.g., across grade levels within a school, then across teachers in a grade level, and so on). This data-driven process leads to concrete actions that teachers and administrators can take to improve schools.

One professional development offering described by Pioneer RESA staff involves acquainting principals with their school's data trends. At the request of a school system, the agency will convene a group of principals and have them log into the *Insight* Data Framework. Then Pioneer RESA staff will guide them through reports and be available to answer the principals' questions. When the principals are comfortable in the system—and generally this takes less than an hour—the staff of the service agency will lead discussions about data trends. According to Jeremy Williams, "drawing conclusions and looking for trends is where *Insight* is really helpful." Or, as

a theory of action, "school improvement tools lead to improved organizational and instructional practices by identifying trends specific to student populations, class periods, domains, or performance levels." The *Insight* Data Framework produces the kind of information that school and system leaders can use to identify professional development and other support needs. The growth model can also suggest organizational improvements, such as modifications to the schedule or promising collaborations between teachers.

The model is easy to understand, and its simplicity is another advantage in the school improvement context. Decisions based on complicated or unintuitive measures may inhibit buy-in and motivation. The Pioneer RESA staff characterize *Insight* as a direct, "no excuses" lens into student growth patterns, and the transparency of the model gives it power in school improvement discussions.

*Validity for teacher assignment and evaluation*

Traditionally, it has been difficult for school system administrators to know whether their systems are assigning teachers efficiently. For example, some teachers may be more effective with low- or high-achieving students, and some teachers may be more effective earlier or later in the day. The *Insight* growth reports may be able to highlight precisely this type of information. Demonstrations by Pioneer RESA staff suggest that the model can extract usable information about efficient teacher assignment from the student performance data.

Regarding teacher evaluation, according to the Pioneer RESA staff, subjective evaluations of teacher effectiveness may not align with results from the growth model. However, teacher effectiveness is currently undergoing a thorough reexamination. The research on teacher effectiveness is rapidly growing and changing, as a host of new research funders (especially philanthropic organizations) and a spate of new technology are entering this landscape. Perhaps most significantly, the Obama administration's Race to the Top competition rewarded applications whose teacher evaluation systems reflected student performance data. Furthermore, a 2009 report from The New Teacher Project found that 98 percent of teachers in 14 large school districts were evaluated as "satisfactory" (Weisberg, et al., 2009), a strong indictment of evaluation systems based on observation.

The *Insight* growth model provides a context for defining effective teaching analytically through results. The focus, for each teacher, is the average change in their students' percent correct, adjusted by the average performance of all students in the reference population. Each student's percent correct, so adjusted, can be thought of as a "residual percent correct." *Insight* labels teachers as Highly Effective, Effective w/ Gain, Effective w/ Loss, and Ineffective, based on the statistical significance of their growth model results, the change in residual percent correct averaged over the teacher's students.

These four classifications and the rules to determine them have not been formally tested, and these labels carry powerful messages, so caution in their application is warranted. However, Pioneer RESA staff report that the distribution of teachers to these classifications has been informally reviewed by some stakeholders and found to be generally reasonable.

*Validity for school and system accountability*

The *Insight* model can produce data at the subject level. A beginning bar for a brand new school is possible, as long as the new school is composed of students who came from other schools and have a prior year score. In fact, using the Georgia testing ID, students who move across school systems can be included in the model. While these features of *Insight* may support its use for system and school accountability, the model has other features that may not agree with these aims. For example, students who do not have scores in successive years do not show up in *Insight's* growth calculation; accountability measures generally include all or nearly all students.

As was the case with both school improvement and teacher assignment and evaluation, the *Insight* growth model can produce accountability measures that have validity on their face. The common cause for this face validity across contexts is that the basic building block of the *Insight* growth model has salience: the change in a student's percent correct relative to the mean (or the change in a student's residual percent correct).

**Caveats and threats to validity**

The *Insight* growth model holds a great deal of promise, particularly for school improvement, but there are elements of the model where cautious interpretation is warranted. Notably, the content covered in the CRCTs varies from grade to grade, so that domain scores produced by the model are based upon weighting. A domain label may represent a reasonable summary of what is measured by a domain score, but the domain categories are not as clean as they would be if the content mix did not vary. It should be noted that similar models based on the CRCT data across grades would encounter the same challenge.

As mentioned above, the teacher effectiveness labels are based upon whether there was a statistically significant change in the residual percent correct, averaged over all of a teacher's students. Additional work is needed to determine if those labels are justified. Furthermore, there are other ways that a label of "highly effective," for example, might be applied to a teacher, such as multiple years of meeting a certain standard.

The greater the stakes of judgments and decisions coming out of the model, the stronger the need for additional validation work on the model. In particular, correlations to other measures and

formal review of the model with stakeholders are strongly recommended if the model's scope or consequences will increase.

Apart from any issues of how the growth metric is calculated, a caveat related to staff capacity seems in order. From discussions with the Pioneer RESA staff, it was determined that some of the most useful reports coming out of *Insight* are the products of special requests. It is not clear whether there is a way to automate more of the reports and lessen the need for special requests. Apparently, some potentially important reports are labor intensive. If the stakes of the reports increase and more reports are requested without increases to staff capacity, there is increased risk of error, and the validity of the process is compromised.

## Comparison with the student growth percentile model

Arizona, Colorado, Indiana, Massachusetts, and Nevada, are exploring or implementing some version of the student growth percentile model developed by Damien Betebenner (2009). The *Insight* matched cohort comparison model shares some key features with Betebenner's model. In both models, there is not a clear interpretation of proficient performance; rather, in each case the model is normative. Information about proficiency is presumed to be gathered from other information or reporting outside of the model. In addition, student growth percentiles and the *Insight* matched cohort comparisons provide a normative context for describing student growth. In both models, judgments of whether growth is sufficient are resolved by external criteria. *Insight* uses the statistical significance criterion to apply teacher effectiveness labels, for example.

## Conclusion

The *Insight* growth model is technically sound. Its greatest rhetorical power comes from its easily understood calculation, which fundamentally makes sense to teachers and administrators, according to staff of Pioneer RESA. Further study can document this directly. Do the model and its calculation make sense to other stakeholders? Again, further study can explore this. The developers have been transparent about the model's limitations and have emphasized that it is appropriate to school improvement discussions. Some systems are using it as part of teacher evaluation. Pioneer RESA has been attentive to how systems are using the model in this context, ensuring that other evidence is always present in teacher evaluation discussions. This monitoring is appropriate and necessary: the model produces information that can be part of a teacher evaluation conversation, so long as other evidence is also present. As with other normative models, the *Insight* model does not speak directly to pegging student proficiency, so this raises a question, what does a unit in this model (a percentage point in percent correct) mean in the context of student proficiency? The model produces labels related to teacher effectiveness that

may or may not be on target—more information is needed. The model is useful in the school improvement context, and it is taking root in the teacher evaluation context because educators can see in the output of the model differences across teachers. (Though it is less clear whether the magnitudes in the output are universally and commonly understood.) For the model to take root as a measure of accountability, educators and policy makers will need to recognize in the model valued outcomes that can be understood as units of school and system performance.

**References**

Betebenner, D. (2009). Growth, Standards, and Accountability. Retrieved December 14, 2010, from http://www.nciea.org/publications/growthandStandard_DB09.pdf.

Jamentz, K. (2001). Accountability dialogues: school communities creating demand from within. San Francisco: WestEd Publications.

Pioneer RESA. (2011). Insight: a growth model for Georgia. (April 19, 2011 version). Cleveland, GA: Author.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: our national failure to acknowledge and act on teacher differences. Brooklyn, NY: The New Teacher Project.