

Growth Models ISBE Meeting-July 26, 2010

10:00-3:00

Present at Meeting: Lizanne De Stefano, Chair; Kathlene Shank, recording; Amy Alsop; Carmen Acevedo; Ellen Cwick; Joseph Matula; Debbie Meisner-Bertauski; Robin Ehrhart; Mark Doan; John Byrne; Robin Ehrhart; Brian Durham; Linda Tomlinson, ISBE; Laura Cresap; Joyce Zurkowski, ISBE; Colleen Legge; Amy Nowell; Travis McQuire; Elizabeth Freeman; Sue Walter;

Item #1: Website

www.isbe.net/gmwg is the site access; readings and schedule of the meetings are available on the site. Meeting summary will also be on the site; these will not be formal minutes as such.

Item #2: Meeting Topics

Chair DeStefano shared that today we will discuss the Tennessee Model; the Chicago Value-Added Model; and the Colorado model.

In August we will look at the longitudinal data system; we will also talk about what an Illinois model should have in it.

Item #3 PEAC

Joseph Matula is on this Growth Model group and PEAC.

July 16th PEAC met; heard information on a survey that is in progress. The Superintendent from Evanston talked about how they are using tests in Evanston Skokie #65. The Charlotte Danielson Model and Learning Points also gave presentations. Another survey system was also shared that is used to evaluate principals.

A more thorough summary of this meeting is available on the ISBE website for this group. August 20th is the next meeting. The focus of this group is on principal and teacher evaluation. 50% of teacher and principal evaluation must be based on "student growth". The state designed system will be used if a district and the teacher representative group (union) can not agree on a system then they will default to the state designed model.

One question asked was "what will happen with districts that have invested considerable resources in data models?" Chair DeStefano said she would be sure to ask the ISBE person sharing the longitudinal data system at the next meeting to address this question.

Members shared what their districts were doing relative to common core standards and components of evaluation systems. One member shared that we need to remember that districts across the state are in varying stages with some who have not had the resources to do as much as other districts have. A question was asked about districts continuing to align to the common core standards and state standards. The unofficial answer is to continue to align with the common core that is already done.

Another question was about having a timeline that showed the areas we would have common core in and when. Linda Tomlinson explained that we already have numerous dates pushing this work that must be met (See summary for June 16, 2010).

Chair DeStefano suggested it may be helpful if ISBE would provide information to districts on the common core standards and where the State is going so districts could plan and move forward so that work is not done that would ultimately be undone.

Linda Tomlinson suggested that it would be good to look at assessments and how these relate to principal and teacher evaluation. It will be necessary to have assessments that are reliable and valid that are aligned to the common core standards. Valid assessments are a key element.

Chair DeStefano shared her personal professional thought that districts will find it useful to continue to discuss the common core and alignment to it. That the common core will be around for a significant time; the ones that have been developed will not go away any time soon. Focusing on the core will provide a framework.

Concern about teachers' being fully involved and feel like the work they have done is valued was shared. The statement was made that it is important that teachers can not feel disenfranchised.

Debbie Meisner-Bertauski shared that higher education will be bringing a group together to discuss what we need to do to bring post-secondary and high school professionals together to discuss how to move forward with curriculum mapping to make it more seamless between higher education and secondary education.

Item # 4: Tennessee Growth Model (TVAAS)

Chair DeStefano shared about this model. One reason Tennessee was picked was because it is one of the first systems approved, 2006. It is probably one of the most mature; has been used since 1998 with implementation beginning in 1992. It is a value-added model. Tennessee was one of the first schools to be given a "race to the top" grant.

Value added is a subset of how we can represent growth. "Status" is the way we currently represent growth in Illinois. (AYP is a status approach) "Improvement" looks at are students

doing better than the group last year; looking at cohorts so it is a way of looking at change. Status and Improvement approaches are not really growth models.

True growth models look at how much on average did students change; it looks at the same group and must be longitudinal.

Value-added is a form of growth model; it predicts where students should be if they continue on the same projectory. If the student exceeds the prediction this is the "value added".

The Illinois legislation does not same we must have a value-added model it says we must look at growth.

The Tennessee growth model (TVAAS) system uses value added analysis as a means of a meaningful way to calculate the influence a district and school have on the academic progress rates of their students. Chair DeStefano says this is a "giant" which produces significant amounts of data. Tennessee is known as keeping as many students in as possible in the system; it is known to be very inclusive. Three years of student data is used when looking at a teacher's performance; five years of data is used to look at a given student's performance. Students who take the alternate assessment are not included in the system. Tennessee has ways to look at different forms of delivery such as team teaching, resource room, departmentalization, etc.

Tennessee does not adjust for socio-economic, race, or other factors. It controls for prior achievement. DOE likes that there are no factors considered.

The strength in this system is that it is multivariate, longitudinal analyses.

Teachers must verify annually that the students attributed to them in the system are the ones they are responsible for; the students must also be in the school for a minimum amount of time.

This system requires, in Dr. DeStefano's professional judgment, a "vertical scale". The Tennessee tests have been stable and in place for a considerable length of time. It has been used in grades 3-8; it has been used in math, reading, science, and social sciences in Tennessee. (See page 19, Guide to United States Department of Education Growth Model Pilot Program 2005-2008,CCSSO.) It takes a considerable amount of psycho-metric expertise; Tennessee has really invested itself in psycho-metric expertise.

The Tennessee system does not share how data is analyzed; it is a very traditional value added model. The software is "proprietary". Tennessee has a very strong statistical foundation.

Data-driven conversations have as a foundation 15 years of data. The system can drill down to a group of 45 or more. Teacher effect reports are generated. Individual student achievement data is also available; the school can give this to parents but parents can not access the data on their own.

Tests to work in this growth model must be “broad”; it must be sensitive at both ends of the continuum. Meaning the standard error of measurement is the same for low achievers and high achievers.

Under AYP “status” has been used but since 2006 “growth models” have also been useable. Three years of data must be used if “growth” is used. Early returns from seven states that used growth models under the NCLB pilot program indicate that growth models produce little to no difference in AYP ratings using growth model from the status models (see page 6, Implementer’s Guide to Growth Models, CCSSO). In Tennessee 7 schools did better with growth measure over status measure of AYP. These 7 schools tended to have lower performing students.

Dr. DeStefano shared that the powerful aspect is the availability of data overtime.

The data is based on “projection” of student performance and not actual performance of students. Three years of data is used. Gain of a school, teacher, or student is measured against state averages. It is linked to high stakes decisions such as “Hope” scholarship and “teacher evaluations”.

Joyce will follow-up relative to three questions: mobility of students and use of the student test scores in evaluation of special education teachers. Another question would be that 8% of teacher evaluation was based on student growth prior to “Race to the Top”, what is the percentage now.

There are aspects of the Tennessee model that do not match Illinois especially the vertical nature of the tests used. One assumption of the Tennessee model is that students are randomly assigned to teachers. This is a piece to keep in mind. It is not a super sensitive system. It is also based on “projection” rather than actual growth. It is a statistical model. There is a 33% swing between groups of students relative to exceeds, meets, does not meet.

The Tennessee model costs \$2-3.00 per student.

Slides provided by Tennessee and shared by Dr. DeStefano will be on the group website.

Item #5 Chicago Value-added

This presentation was made by Amy Nowell.

This model is a value-added model. WCER (Wisconsin Center for Educational Research) is working with CPS; the focus has been on whether schools are performing or are effective. The slides from this presentation are also on the website for this group. Conditions are controlled for in this model. Real data is used and based on conditions that can be controlled. It looks at adjusted growth. The assumption is that what remains after controlling for conditions beyond control what is left is the "growth". ISAT is being used; it assumes it is a vertical scale test. Chicago control factors in the model are students' prior performance; free/reduced lunch status; ELL; IEP/Special Education status; (primary disability); gender and this year race is being added. It is used in principal evaluation but not teacher evaluation. There is a swing between groups of students that adds to the standard error of measurement that makes this less useful for teacher evaluation.

This model uses grades 4-8 with grade 4 being the baseline with performance measured by the ISAT. The value added is the average student gain on ISAT over and above the district average over 3 years.

It is a multi-regression model; it is "means" based. It makes "correction for pretest measurement error". It also corrects for "shrinkage" which relates to school size. The bias is that there is higher growth for higher achieving schools.

Every year data uses that year's data; there is no projection. It is not based on multiple years of data; it is based on one year. Pre-test is always Spring; the pre-test is the spring ISAT.

This model controls for pre-test measurement error and student's changing schools (dosage).

This is a classroom level model; there are data issues linking students to teachers. There is also difficulty with multiple teacher effects such as special education resource rooms. Data is an overall school and grade level value added in reading and math. Data does provide that CPS can differentiate between top 10% schools and lowest 10% schools but rest are in the middle.

Value added is used in "performance policy". It is used as 25% in the principal evaluation. Trend is used in the principal evaluation. CPS will need to move to use it in teacher evaluation.

CPS concerns include "stability of value added". Movement between categories wasn't as great as once thought might be true.

"Growth" was not part of the CPS focus whereas in Tennessee it was; also, Tennessee is based on "projection" whereas CPS uses "real" data. CPS has determined what it will control for

whereas Tennessee does not control for any factors. Student is at the heart of the Tennessee model.

Item # 6: The Colorado Model

This presentation was made by a representative of the Colorado Department of Education, Bill Bonk.

Joyce shared that the Colorado model is being used by 12-15 states; not all states are using it in the same way but rather in varying ways. This conversation is just in the initial stages.

The basis of the growth model is a “normative comparison”. Test scores in themselves have no real meaning. The point of the Colorado growth model is that data must be considered in a human context. Test scores should not be viewed as precise. The best way to understand scores is not in terms of a vertical scale but rather to look at change in scores and what this means for learners that started in differing places. The model uses the current test score as the dependent variable. It is a regression growth model (Additional information, see page 15-16 of the CCSSO document, Guide to United States Department of Education Growth Model Pilot Program 2005-2008).

It is a norm model and the scores change every year. It does not need or assume vertical scores. This model is based on the set of previous scores and looks at where the new score fits. It is expressed as a percentile score. It gives you “growth percentiles”. As it is a norm model there will always be the lower percentages and the higher percentages. Colorado is using its Colorado tests across grades 3-10 in reading, math, and writing plus the Colorado ACT at grade 10.

Current growth is seen as a leading indicator. A goal is that this data is communicable to stakeholders. How to visually represent data has been a very important consideration and commitment of time in Colorado. The system has been explained on a website which is being re-launched this fall. Colorado’s goal is to have the data be transparent at all levels and understandable at all levels.

The essence is to look at test scores in one year and compare to one, two, and three years later; only scaled scores are used. It compares the score with other students who started in the same place. Scores are only compared to others who started at the same point across the state. Two years data across the state are used to arrive at average percentiles then three years is used to arrive at average percentiles. For each individual student the maximum number of years of data

is used (could be for each grade 3 through 10). A quantile regression is used with the expectation that all students will reach proficiency within 3 years of by 10th grade.

Colorado is not using a value-added model; no demographics are considered. It does not know student attributes or which school. It is normative at its base and produces individual interpretations. This is not a cohort model; the group is mathematically constructed.

The model produces individual interpretations. Compares growth to a peer group that was mathematically determined; 50th percentile is the norm. 50% is not the target; it is not the standard. Median is used not "means". Medians must be used as it is percentile based. Medians are used to account for differences in intervals.

This model is used at the district and school level. Colorado does not currently have an application of this model to teacher evaluation.

Data is displayed and made available to stakeholders. There are two versions. School and district growth summary reports are provided. There is a "SchoolView.org" website; the protected level for number is 20. Data in the new version will go back to 2003.

The model was approved by the DOE in the growth model pilot program process but Colorado elected not to use it in place of AYP. Colorado is currently using the status AYP measurement for NCLB not this growth model.

Although factors are not controlled for group data can be broken out by school, grades, race, etc.

A question asked of Dr. Bonk was "capacity" necessary to do the work he is describing. Dr. Bonk responded that they have had multiple sources of funding, including private. He said there is an open source called "R" where Illinois could run its data. Oracle itself costs \$1 million. If a State signs an article of understanding with the Colorado State Department of Education the State can use what has been developed by Colorado for free.

Professional development has been focused at the district level not the teacher level.

Data for spring testing is reported by mid-August.

Colorado will be introducing the concept of "adequate growth". This is based on "catch up and keep-up growth". Since it is norm referenced there is always 50% growth. "Catch-up Growth" refers to crossing into proficiency; "Keeping up" represents maintenance. The performance framework looks at these factors; this framework is being used for districts, schools, and varying groups (free and reduced lunch, ELL, special education excluding those being test through alternative assessments, plus two others). The "growth is normed" not the scores.

Dr. Bonk knew of one district using the data as “one factor” in the teacher evaluation process. Colorado as a state uses this information as description information. The data is not used as causal information. It is not viewed as precise and as a singular information point but rather one place to begin the conversation.

The metric is “growth” not test scores.

In Colorado discussion of growth models began at the district level in 2001; the State adopted the model in 2007. It became a topic of research for the Colorado Department of Education in 2008.

Dr. DeStefano reflected that this model is complex.

THE NEXT MEETING of this group is AUGUST 30th in Springfield, Il. at 10:00 a.m..