



UNITED STATES DEPARTMENT OF EDUCATION

OFFICE OF ELEMENTARY AND SECONDARY EDUCATION

THE ASSISTANT SECRETARY

APR 5 2007

The Honorable Patricia Hamamoto
Superintendent of Education
Hawaii Department of Education
1390 Miller Street, #307
Honolulu, Hawaii 96813

Dear Superintendent Hamamoto:

Thank you for submitting a proposal for the U.S. Department of Education's (Department) growth-based accountability model pilot project. I appreciate the work you and your staff have done to participate in this effort so far. The Department continues to believe that this pilot project can help determine whether growth models will provide a fair, reliable, and innovative mechanism for holding schools accountable for ensuring that all students reach grade-level proficiency in reading and mathematics by 2013–14.

As you know, the panel of peer experts reviewed Hawaii's growth model on March 15-16, 2007. During this review, the peers raised a number of substantive concerns with the structure of Hawaii's model, indicating the model was not acceptable for implementation in the 2006–07 school year. I am enclosing a copy of the peer report for your consideration. Based on the significance of the peers' concerns, the Department has decided not to approve Hawaii's proposal for implementation in the 2006–07 school year. I anticipate, however, that there will be other opportunities for Hawaii to implement a growth model in the future. As noted in the *Building on Results: A Blueprint for Strengthening the No Child Left Behind Act*, the Department's reauthorization proposal would permit States to include a growth model to measure adequate yearly progress, provided the model is approved by the Department. I urge you to consider carefully the peer reviewers' feedback as you work to refine your growth model for the future. My staff and I are willing to discuss the peer's concerns with you to help refine your model.

The peers identified several strengths in Hawaii's proposal, noting the creative and technically advanced aspects of the model. The peers commended Hawaii's inclusion of all students, not just those who are below proficient, in the growth model. The model estimates the probability that a student will become proficient and aggregates this information to form accountability determinations for schools and subgroups. The peers also noted Hawaii's proposed model uses sophisticated methods that can handle issues of missing data and develop unbiased and precise estimated standard errors. In addition, the model reassesses probabilities at different time periods.

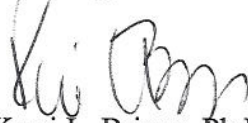
However, the peers noted specific concerns with fundamental components of Hawaii's proposed model. The peers were unclear how the probabilities for students predicted to

be at the proficient level were aggregated. As a result, the peers were concerned that the expected performance of high achieving students would mask the students with lower rates of expected growth.

In addition, the peers were concerned about the State's capacity to implement the proposed model. The peers noted several key components of Hawaii's data and assessment systems necessary to fully implement the proposed model were not well established or had not yet been validated. The peers noted that Hawaii must still create its vertical scale, adopt new academic achievement standards, and that 2006-07 administration will be the first full administration of the new assessment. The peers were also concerned regarding the match rates for certain student groups, noting the lowest match rates were for those student groups with the largest achievement gaps. (Please refer to the enclosed peer report for details.)

Again, I appreciate your interest in the growth model pilot project and your continued efforts to ensure quality education for all children.

Sincerely,



Kerri L. Briggs, Ph.D.
Acting Assistant Secretary

Enclosure

cc: Governor Linda Lingle
Robert McClelland

PEER REVIEW
NCLB GROWTH MODEL PILOT
U. S. Department of Education
March 15-16, 2007

PEER REPORT

State: **Hawaii**

Clarifying Call to State

No call necessary

Questions and responses (please note within each question if the State will provide additional information)

Overall Recommendation

<u>Recommend to Accept</u>	<u>Recommend to Accept with Conditions</u>	<u>Recommend to Resubmit</u>	<u>Recommend not to Accept</u>
0	0	0	14

(Two Peers Abstained)

Comments to support recommendation:

Summary of issues

Strengths

The proposal makes sense because the probabilities are reassessed at different time periods

- 1) The proposal does include all students in the model, not simply those not proficient.
- 2) The proposal is an improvement on growth models simply attempting to count the number of students projected to meet a certain threshold.
- 3) The model estimates the probability an individual student will become proficient (and uses this information to form aggregates for schools and subgroups).
- 4) The proposed model use sophisticated methods that can handle missing data.
- 5) The proposed model uses sophisticated methods to develop unbiased and precise estimated standard errors.

Weakness:

- 1) School effects (there are no random school effects) are not brought into the analysis and students' growth trajectories are shrunk to state rather than school means. This will unfairly advantage very low achieving schools, producing more positive predictions than are likely to actually occur. The prediction validity test will likely fail here.
- 2) Clarify how probabilities (proportion of students predicted to be proficient at the school level) are averaged together, to make sure that high achievers don't mask non-improvements among low achievers in producing the final results used for AYP determination.
- 3) Describe the plans to explain the plan to the public.
- 4) We are concerned that there is missing matching data on some subgroups. Specifically, the lowest match rates are for those subgroups with the largest achievement gaps. If the state proposes to include only those students with at least two test scores, then precisely those students who are behind will be left out of the model. That state should clarify why matching rates for some subgroups are low and how this impacts results.
- 5) (related to 4, above). The state's methodology can clearly include students with a single score. Can the state provide rationale why it might be preferable to only include students with multiple scores – although this has the problem discussed in (4) above.
- 6) The state proposes a creative and technically advanced model that is dependent on several key elements being established and validated in order to fully implement a meaningful growth model. However, the state notes that it still must create a vertical scale, adopt new performance standards, receive final approval for the assessment system, have multiple years of the new assessment in place, equate the previous assessment with the new assessment, and attend to issues related to the alternative assessment. These are significant efforts that cast doubt on the state's ability to fully implement the growth model this year and were the principal grounds on which the proposal was judged not acceptable at this time.
- 7) If the state intends to resubmit in a potential future round of the pilot program the state should also include simulated data that demonstrates explicitly how the model will work for different types of students (attending different types of schools) and how the model works over time following the same set of students over time. Further, the simulations should include how students who are retained or change schools or LEA's are handled. Simulations for students taking the alternative assessment should be presented as well. See also general guidance for future proposals below developed by the

panel.

An additional panel comment:

Since the idea of growth models in the context of AYP decisions under NCLB is relatively new, and often the procedures proposed are complex and may be untested, the panel is concerned about the prediction validity of the system proposed. We suggest that in any future new proposal you do the following:

- 1) Please provide in one place in the proposal all of the necessary rules, procedures, statistical models and estimation procedures (if you are employing a statistical model), AMO information and so on required to operationalize your proposed system. The detail should be sufficient such that an independent third party could, in principle, build a system that reproduces your AYP growth results. It should describe how all children will be handled including: those that change schools; change LEAs; are retained in grade; transition for one school type to another (e.g. elementary to high school); and who participate in alternate assessments or with the use of accommodations. A resubmitted proposal that does not meet this standard will be considered unacceptable.
- 2) Provide multiple illustrations/simulations of how individual students beginning at different test levels and grades, and progressing at different rates over time would be judged under this system. These examples should be chosen to illustrate a diverse range of the data patterns that might occur and should follow these same students over time.
- 3) Additionally, the simulations should include students whose scores may be treated differently under the proposed model (i.e., retained students, students with missing scores, and students who participated in an alternate assessment). Please detail if there are cases where scores for some groups of students might be treated differently in the model or in an alternate growth model.
- 4) Provide an assurance to the panel that you have carefully considered the overall prediction validity of your system. That is, the growth accountability option involves making predictions, based on children's past academic growth, about the likelihood that children will in fact achieve proficiency at some future time point. We are particularly concerned about the possibility that the system might over predict subsequent proficiency rates in very low achieving schools. We would appreciate any information that you can provide that addresses this concern. If some extant prior years' data permit, you could apply the system detailed under 1 above to student results, say in 2004, to make predictions about status attainment, to say in 2006, and then compare the prediction to the actual status attained. An unbiased system would not result in systematic discrepancies between predictions and actual attainments.
- 5) Some members of the panel were particularly concerned about growth model formulas that do not take into account the possibility that some students who may be judged proficient in the current year may regress below that threshold in subsequent years. If growth projections are not made for "currently proficient students" how will your system take this consideration into account? (This consideration is especially relevant to the issue of prediction validity noted under 3 above.)
- 6) Please indicate if there are any sub-groups of students for which you will be unable to apply your prediction model. Note an acceptable growth proposal must detail a strategy that includes predictions for all students tested.

(Use additional space as necessary)

Specific Strengths in the Proposal

Using your notes from the Peer Review Guidance, please note areas where the proposal was especially strong, ingenious, high quality, or exceeded the Peer Review criteria. Please cite specific aspects of the proposal and include references to the Peer Review Guidance criteria (e.g., B.1.2.1) and the proposal (e.g., page numbers).

- Holds every student accountable. Estimates probabilities of students being proficient.
- The proposal makes sense because the probabilities are reassessed at different time periods.
- Hawaii admitted that they made a notation error in an earlier submission.

Dissenting comments:

(Use additional space as necessary)

Specific Weaknesses in the Proposal

Using your notes from the Peer Review Guidance, please note areas where the proposal was unclear, incomplete, or did not meet the Peer Review criteria. Please cite specific aspects of the proposal and include references to the Peer Review Guidance criteria (e.g., B.1.2.1) and the proposal (e.g., page numbers).

Sophisticated use of estimation error, but is error conceived correctly? Hawaii should look at forecast error. This might underestimate error. (Minor issue, but technical issue that one is not comfortable with.) I would like to see additional clarification of the estimation of the forecast error. Forecast error and prediction error are not the same (prediction error is based on the residual that compares estimated to actual scores; the forecast is based on prediction error but carried forward to some time point beyond the data. It should be explicitly identified that the forecast error will increase for each time point beyond the available data (if the model coefficients are unknown)).

Estimation of probability doesn't take into account possibility of backsliding from one year to another.

Estimating effects of individual schools, but ignoring random effects. Forecast is based on individuals but is shrunk to overall growth estimate mean. This does not bring school effect into the analysis. Students in schools with low average are being compared to state averages, not school average.

Can states use confidence intervals for performance estimates? Hawaii might be employing a way around using confidence intervals. A peer liked that part of the model. It allows for descriptions of students just above or below thresholds.

Forecast error will probably stretch out the tails of the distribution.

Averaging probabilities can overestimate proficiency if a teacher has a bunch of kids with high probabilities, they can pull up probabilities of kids with lower scores. We think that the system isn't supposed to let high kids compensate for low performing kids, especially without a school effect. Students with similar proficiencies in different schools won't have their different probabilities (based on school) taken into account.

But, how do you take school effects into account, since school environments can change over time? Most of these proposals estimate future growth based on past performance, without any nesting within schools. This argues that all students across all schools are alike, which ignores impact of schools on students. A peer agrees, but notes it's difficult to build that aspect into the model. The model isn't supposed to allow in other factors, like school effects. If Hawaii doesn't use school effect, Hawaii will overestimate growth of kids in low-performing schools, and underestimate growth of kids in high-performing schools. This could let lower-performing schools off the hook, and it lets Elementary schools off the hook, because proficiency assessments are in 7th grade, even though you're estimating probabilities in elementary schools. Because similar students cluster in schools, characteristics of those kids show up in the estimates. So, there's still, effectively, a demographic effect in the model.

This is a complex model and would be difficult to explain to the public.

There's an issue about to which school you attribute the students' growth? This is probably a nonissue, because once a student hits a low-performing school his or her progression will probably be dampened.

Also, it looks like Hawaii's method hasn't yet been tested, and there might be a host of practical difficulties. There's an issue because the state could have difficulty matching data, so there might be a large amount of missing data on certain types of kids. This could be mitigated by knowing other data.

Is the assessment system safe, secure, and of known high quality? Imposing a sophisticated analysis system on top of that makes the whole system in need of quality control. Gut reaction: nice idea, but it's too soon. This is still just a conceptual model, and it's hard to approve such a model. We can give Hawaii feedback as to how they can do this probability projection. We need to know if it's compensatory or not.

Point 1: There's enough uncertainty that approval is unwarranted at this time. Point 2: It's an innovative model, and it would be nice to see someone try this at some point when the state is ready to do it.

Point 1: We should be clear about how we build school factors into the projections. If we guide Hawaii on this, we should get clarity on the issue ourselves. Point 2: We should give them counsel about how to resubmit in a more clear fashion. There are two issues in clarity. One is how clearly it's conveyed, the other is the simplicity of the underlying model. The underlying methodology can be complex, if it can be presented clearly. And, it's most important for the system to make good projections. But, can we make those judgments without results? You need multiple years of data to know how well the systems work and whether it's valid.

A peer would like to know more about the disparities in the match rates across racial groups. A peer noted it could be because of small groups of racial minorities, but that's just speculation. It would be nice to know the real reasons behind the non-matches. In the future, we would like to know this.

A peer doesn't think that the crediting growth between schools with different structures was done very well. Whatever school the student is in gets credit for the growth. If a middle school has strong feeder elementary schools, the middle school gets the credit. Data shows that the last school the student is in actually has the dominant effect on the student's growth. This model doesn't try to get an estimate of the school effect on student learning.

Point 1: Their standard of growth is where the student should be in 3 years. But, this will probably fail because they're projecting for kids that have lagged in the past.

The consequences are for schools where the student is now.

Dissenting comments:

(Use additional space as necessary)