STATE OF FLORIDA
DEPARTMENT OF EDUCATION
AMERICAN INSTITUTES FOR RESEARCH

FLORIDA'S RACE TO THE TOP
STUDENT GROWTH IMPLEMENTATION
COMMITTEE MEETING

University of Central Florida
Teaching Academy Building
Orlando, Florida

Thursday, May 19, 2011
Volume 1

DEPARTMENT OF EDUCATION:
KATHY HEBDA, Deputy Chancellor for Educator Quality
JUAN COPA, Director, Research & Analysis

AIR MEMBERS PRESENT:
JON COHEN, Ph.D., Executive Vice-President
HAROLD DORAN, Ed.D., AIR, Principal Research Scientist
CHRISTY HOVANETZ
MARY ANN LEMKE

American Court Reporting
850.421.0058
Good morning, everyone, and welcome to the next meeting of the Student Growth Implementation Committee under Race to the Top. I'm Kathy Hebda, I'm the Deputy Chancellor for Educator Quality with the Department of Education. I'd like to not only welcome our committee members this morning but welcome our audience that might be watching over the web, and anybody present in the room. If there are audience members present in the room, we are very, very pleased that you're here. We'll have hard copies of the power point for you at the lunch break, and we also would remind you that this is a webcast and is always open to the public. We are very pleased that folks are interested in this meeting. This is a working group. The committee members, the ones that will be speaking during this time and doing the work of the committee, and we appreciate your participation.

Members, I'm going to let you begin by introducing yourselves one more time. Most of you were here in person at the last meeting, but if you were participating from afar or not quite in the room, I'm going to let you do that. If you would just say your name and your position and who you represent, and we'll start with Ronda.

MS. BOURN: I'm Ronda Bourn. I'm supervisor of special projects at the Northeast Florida Educational Consortium.

MS. EDGECOMB: Doretha Edgecomb, school board member, Hillsborough County Schools.

MS. KRISHNAIYER: Latha Krishnaiyer, Broward County.

MR. MOREHOUSE: Lawrence Morehouse, president of Florida Education Department and professor of USF.

MR. LeTELLIER: John LeTellier, music teacher, Stanton Weirsdale Elementary School, Marion County.

MS. ACOSTA: Sandi Acosta. I'm a middle school science teacher at Kenwood KA Center in Miami.

MS. KEARSCHNER: I'm Linda Kearschner, I'm a business owner and I'm on the board of American Court Reporting.
The first couple of things on that just to get everything set up before I turn it over to our experts, you have two sides that cover the meeting agenda for today and tomorrow. We begin at 8:30 each day and will adjourn at 5:00 today and at least by 5:00 tomorrow, no later than 7:00 tomorrow.

One of the things that I would say about the agenda, you can -- I won't read every line to you; you can see that on your own. Our contractors, the American Institutes for Research that worked with you the last time at the last meeting, what we've asked them to do this time is to work towards recommendations to the commission for a value added model to be used with FCAT data, to provide you not just the results of all the data requests that you made of them at the last meeting but also a method for you to make and work towards that decision. And they developed a process that will help you do that and track the information that you receive from that data analysis and from the results of those data runs, and that will help you as work towards making your American Court Reporting 850.421.0058

decisions over the next two days. So I think that's something that you can look forward to and feel comfortable about because I know that you may be thinking, well, you asked for a lot of information last time and how are you going to receive that information and how will you make decisions about that information. And they have a process to help you do that.

So as I mentioned, the purpose of this committee -- just to remind everyone and everybody who might be watching -- the purpose of this committee is to make recommendations and the June 1st recommendations that you make, those that go to the Commissioner, you need to make a final selection by June 1st. That's really the first order of business that you have. It's not the last. June 1st is just a beginning. It's a process, the beginning of a process. It's a very important step, no doubt, but it is a first step and you'll be working throughout the remaining years of the grant to improve the models and recommendations that you make, look at how those models can be communicated, lots of things like that that you have on your agenda for the next three years.

So here you are. You may remember these slides from last time. This is the process for sort of the year one of the grant with relationship to student growth. You already identified the initial models that you wanted and selected models for comparison. That's what you did at your last meeting and determining -- you had a discussion of the variables and business rules. One of the things we'll do after I finish these introductory slides is we're actually going to spend a couple of minutes reminding you of what those decisions were. In case you don't have your notes with you or any of those sorts of things, we'll lay out very succinctly where you've come so far. That will help remind you where you are and what you're going to do next, and of course, anybody who's watching the proceedings, we'll also help them know exactly what's happened if they miss the first meeting.

So you finished all that at the last meeting. What's in the loop here is what you're going to do today and tomorrow. You're going to evaluate those selected models, the results of those things, how they run data using those models and the variables, and the business rules that you discussed; and you're going to compare those things. Remember, I told you we have a process that will help you the next two days to do that comparison in a way that you can feel comfortable about, there will be times for you to stop, reflect, always opportunities for questions and clarifications as you go throughout that process the next few days. Then after the final selection is made by the Commissioner on June 1st then there will be reporting of results. School districts will get their data that was used to evaluate the model, the final model selected, data they can use to make local decisions about how they're going to use it in evaluation systems and their scales for ranking folks or rating individuals making American Court Reporting 850.421.0058
decisions about performance next year. That's
the process they'll be going through this summer
after that data is provided to them.
So that's where you are in your timeline.
Again, this is the goal of the meeting.
It's a simple goal; it's an important goal.
Just reminding you again that your goal is to
make recommendation. The Commissioner does have
the responsibility to make the final selection.
Then, of course, as that model is implemented
next year and the following year and the year
after that, every year there's a process built
into the grant to make sure we analyze how
effective the model was and ways to make
improvements in the model as we go.
One other reminder before I turn it over to
our AIR partners is their role throughout this
process, just like you saw in the very first
face-to-face meeting you had, their rule is not
to make a recommendation. That's your role.
Their role is to fulfill your request for data,
provide information, answer your questions, lend
expertise to the process, but the decision are,
in fact, yours. I want to make sure that
everybody is very, very clear about that.
American Court Reporting
850.421.0058

Remember, I told you we're going to walk
through the background of the decisions made so
far and I want to bring everybody up to date,
I'm going to call on Dr. Hovaneltz to please do
that part for us.
DR. HOVANETZ: Good morning. Just a brief
reminder of where we've been and how we got to
where we are today to look at the results that
we have. A mere six weeks ago we were sitting
here all together and we had our first
face-to-face meeting where we actually narrowed
down to have AIR look at three selected
models, a covariat model with fixed and random
effects and a sustained differences model. We
selected variables that we wanted them to
include in the model and we selected business
rules to guide the evaluation as well. We
looked at and pulled out of Senate Bill 736 the
three specific variables -- students with
disabilities, English language learners, and
attendance -- and defined each of those
variables for inclusion in the evaluation.
We also had a webinar on April 14th where
we did additional identification of variables to
be included in the analysis. So we identified
American Court Reporting
850.421.0058

here I was really sick; this time I feel great.
This time Jon is actually the sick one, so the
coughing is coming from him not me.
This has been for us at AIR in working with
DOE a really exciting six weeks. There have
been a number of things that we really have an
opportunity to look at, think about, analyze
that we're going to bring to you today. So I
think this is exciting.
I'd also like to say because I've been
doing value added for a while and it's my
opinion, which I think is a pretty substantiated
opinion, that the work that this group is doing
is the most comprehensive value added for
comparison and in-depth analysis of different
models that I know of that has ever happened to
go into operation, at least people who have done
the studies, and they thought about value added
models. The depth of what this group has done
has been very, very substantial and we're going
to present a very substantial amount of results
today.
Before I get started, I wanted to stop real
quickly and look back at where we were and see
if anybody had any questions that were pressing
American Court Reporting
850.421.0058
One of the commitments that we have in terms of implementing the value added models and coming back to this group is we want to be clear about what we call an estimand. What are the different value added models actually estimating in clear terms, because these are statistical models that are doing something that has to be understood in terms of their transparency, in terms of what they are estimating.

Essentially, there are two types of models that we’re going to present today. We reviewed three genres of models last -- two genres of models, what we called the layered and persistence model or the learning path models and the covariate models. So now let’s talk about what are the models actually estimating, the differences model is -- we say this, expect students who score the same in the pattern here to score the same and to continue to score the same, and assumes the same amount of growth for American Court Reporting
850.421.0058

Each student in each achievement model. So students within performance level one are expected to have the same level of growth that all other students in that performance level. Students in the third performance level have a similar expectation. I’ll show you a graphic in just a moment. The covariate adjustment models, these models expect students who score the same in prior years to score the same the next year. Expected growth may vary within achievement level. It’s that last part that makes these two models different. In the first model, students within a similar performance category have similar growth expectations. In the latter, students in a similar performance category can have different growth expectations. That is the key difference between these two models.

Let’s look at a visual display. Sometimes interpretation of what the models are actually estimating and some others like to look at some of the plots. I’m going to try to present information in multiple ways today so that it American Court Reporting
850.421.0058

The green line is the line that would be fit through a covariate adjustment model, and essentially, one of the things that we see here in this red line is the simple differences model. It essentially has a slope of one within each of the performance categories, whereas the other model has a slope of whatever it is. It’s not necessarily one; it doesn’t have to be constrained to be one. But it’s not the same within a performance category where you see the red line expects lower growth for students at the lower end of the score distribution relative to the covariate adjustment model. The important point that we want this group to recognize is that the models differ in terms of their expectations for growth for students within a particular performance model. My team want to add anything?

DR. COHEN: Yes, I want to add a little bit. This is -- we’ve boiled it down to just two classes of models earlier.

DR. DORAN: The web books have -- the web books have these, is that correct.

American Court Reporting
850.421.0058

Does everyone understand what the green line is?

The red line, it’s just a simple difference. All we did was we said we’re going American Court Reporting
850.421.0058
to look at your score and subtract out the average growth from within that group. We'll just look at the average growth for kids who started at the same achievement level; and instead of having a best fitting slope, it has a slope of one. You look, it's hard for you to see in some places where these lines differ. But you do see that they cross a number of times. Where the red line is above the green line, that means that you're predicting higher performance among the kids than is typically observed. The green line is what's typically observed; it's also the line hit by the covariate adjustment model.

So down here the difference in models predicting less growth than is typically observed. That would tend to say that teachers teaching these students would typically exceed that more readily. At the other end you see this is where it goes if you follow this, at the end of this group, teachers teaching these kids would have a harder time exceeding that because you're expecting more growth than is typically observed. Does that make sense? That's the difference between the covariate adjustment

American Court Reporting
850.421.0058

Model 1A is the same as Model 1 except that
American Court Reporting
850.421.0058

we have two years of prior achievement data.
What I mean by that is in this regression we have what's called a dependent variable -- that's the current year's score -- and then we use two prior years of achievement data. We call those lags. So one lag means in the regression model, one of our independent variables is a prior test score. If we had two lags then we're using two prior test scores.

The rationale for that and whether or not that matters is going to become clear as we look at some of the results comparing Model 1 and Model 1A.

Model 2. It's the same as Model 1 but estimated with fixed effects. Let me say something now and we'll talk more about this throughout the day if we need to. We estimated the model with fixed effects just as we said we would. We're not presenting them today although we can fully talk about them. Let me explain why.

Algebraically, mathematically, we know that the random effects and the fixed effects are guaranteed to be the same as the number of kids in a class gets larger. That's the constraint

American Court Reporting
850.421.0058

that has to be in place. They can be the same and they estimate the same quantity. Early on in the analysis we saw that the models with fixed effects were yielding unstable results. We spent quite a bit of time looking into this and we know why. There is a business rule -- there were a couple of issues. There is a rule that allows for students, I forget exactly how to say this -- Christy or Jon, remind me -- students who are taught by multiple teachers to be partitioned across these multiple teachers. Essentially what was happening is there's teacher one who has a group of kids, teacher two who has the exact same group of kids, okay? Two different teachers, exact same group of kids.

In a fixed effect model you can't have that, all right? The term we use is called co-linearity. We have to remove one of those teachers. We're estimating the same exact thing. In fact, you can remove one of those teachers because they're duplicated without any consequence on model estimation.

There's another condition happening in the data which causes near co-linearity. They're not exactly the same kids but they're pretty
American Court Reporting
850.421.0058
close to it. It would almost have the exact
same number of kids but they might be off by one
to two. That causes for a problem in the
human estimation of the model. What was
happening was that those teacher effects were
causing -- those teachers in the data, without
conditions it appears to be true -- were causing
for the results to be unstable. In fact, if we
remove some of those issues and we estimate the
model with fixed effects and we correlate it
with the random effects, they are correlated
better than 0.9. A correlation ranges from –1
to 1; a correlation of 1 means there is a
perfect relationship between the two, a 1 to 1
correspondence. A correlation of 0 means
there is no correspondence between the two
estimates. The closer you get to 1, the greater
that correspondence between those two is. A
correlation better than 0.9 tells us what we
hypothesize about this. When we were here last
time six weeks ago that the models were
estimating the same thing and that turns out to
be true in the data.

The issue here is and the reason we're not
presenting it is because there is a business
American Court Reporting
850.421.0058

rule that prevents these models from being
estimated and presented in a reasonable way to
this group, but we do know that we could rely on
the fact that both algebraically and in the real
world the results are highly correlated when we
remove that particular issue and when we don't
take it into consideration.

Do you want to add anything?

DR. COHEN: No, that's basically it. You
wind up having to toss out a lot of teachers
because of fixed effects model; it's not true
with the random effects model. You don't want
to really toss out a lot of the teachers, you
don't really want unstable estimates.

DR. DORAN: All right. Model 3 and
variance. A three-level model that includes
teacher and school effects. Control for two
prior achievement -- two years of prior
achievement, two lags, and varies as to which
variables are included. That's the general
class of Model 3. So we have teacher and school
effects. It's a three-level model because the
data are structured in this -- we have students,
teachers, and schools. That's the terminology
that we use. We have two-level models and the
American Court Reporting
850.421.0058

and that three-level model.

Model 3A has no additional variance.

That's it. The description that we see.

Then we have Model 3A1. Model 3A1 is Model
3A but it differs only in terms of the number of
prior achievement test scores. It has one lag,
not two.

Model 3B. This is the description, but it
includes ELL, SWD, and attendance. We use two
prior test scores -- oh, there's a note there.

It always uses two prior test scores.

Model 3c, ELL, SWD, attendance, and the
following additional variables -- class size,
homogeneity of class composition. Let me
explain that variable. That is a variable that
describes how similar students are within a
class, all right. So we construct a variable,
call it the homogeneity variable, and it's
essentially -- I think we've got another slide
that describes it, but I'm going to mention this
now. I have to find the easel to talk about
things multiple times. It is the difference
between the students at the 75th percentile
within a class and the 25th percentile within a
class. So if students within a class are very
American Court Reporting
850.421.0058

similar in terms of their prior test scores,
that difference will be small.

If students within a class, if a teacher
has a class that is very different in terms of
their ability, that number will be very big.

It's a spread. It's essentially a spread
between kids within class. How different are
kids within a class.

Mobility, student mobility, and difference
from modal H. Christy will describe those in
just a moment when we get to those slides. And
we use one or two years of prior achievement
data -- here again, this one uses two prior
achievement scores, okay.

And last, Model 4 is the differences model.

All right. This is the language description,
the narrative description of what these models
are doing. In your folders you have this -- can
everybody pull this out to make sure we all have
this here? I'm going to refer to this as the
scorecard throughout the day, today and
tomorrow.

This is the same as what you see here, but
you're going to want to keep this next to you
throughout the day. We're going to talk about
American Court Reporting
850.421.0058
the models and all of our slides use labels like 1, 1A, 3A, 3A1, and so forth, and we don't want you to get lost in the details of the models.

So what we've created is this little scorecard; and essentially what you see is Model 1 and its key characteristics. So, for example, Model 1 has one lag and had teacher effects only. No school effects. The covariates that are included are SWD, ELL, attendance, and the effects, the teacher effects are random, and school effects if they are included. There's no school effects in Model 1.

All right. Now I'm calling this a scorecard because one of the things that we're hoping we will do is by the end of the day we have to make a recommendation, and Sam is going to facilitate a conversation. What we have here in terms of notes, these are the primary evaluation criteria by which we'll be looking at these models today. We have data that show you how these models stack up against each other based on those criteria.

One of the things we're hoping this scorecard will be used for is as we talk about precision, for example, and the precision of the American Court Reporting 850.421.0058

different models, you will come up with your own ranking system and maybe you'll use a 1 to 10 scale. Maybe you'll use a 1 to 50 scale. Maybe you like happy faces, pluses, minuses, whatever is comfortable for you, and if you like a model, for example, and you're using -- you might put for that particular model, Model 1, a precision plus or an 'A'; or if you don't like it, you might put an 'F' based on the data and the results. And essentially you might use that to have a ranking for each model on your own personal scale for precision. Then include school effects. Do you like to include school effects? And so forth.

By the time we get to Sam and he facilitates the conversation about whether or not you're ready to make a decision or intend to make a decision on a model, you'll be able to look at this and you'll look across the rows, and you'll say, well, I look at Model 1 and I have all the sad faces. They're all minuses there. In my own opinion in these criteria, I can eliminate that model from my choice. I look at Model 3B. I have all those pluses there.

Then your decision by the time you get to the American Court Reporting 850.421.0058

1 conversation with Sam is not random.
2 Have you forgotten anything? You've evaluated the models based on empirical criteria. There's going to be a ton of discussion and other things that you care about perhaps beyond this, but the goal here is we have to have a process by which we evaluate the models that's better than our opinions on what we think the world should look like. So we made an attempt not only to estimate a large number of models with multiple areas, but to also come up with what we believe are reasonable indicators that you can use as the lens by which you can evaluate the model, then this decision is yours. It's not us standing here telling you this model needs to compute, it's statistically very nice. It looks good, the plot looks good, and we remove hopefully all of it. Keep this by your side throughout the day.

Jon, go ahead.

DR. COHEN: I just want to help with a little bit of organization here. It's not entirely clear from the chart -- I mean, it's there but it doesn't just pop out at you --

Model 1 doesn't include school effects. It's

American Court Reporting 850.421.0058

do you include? And we have everything that

American Court Reporting 850.421.0058
ranges from none to a few to the kitchen sink.

MR. TOMEI: Quick question. I'm looking at the chart and it appears that the lags for 3A and 3A1 may be inverted, based on what's on the slide; is that correct?

DR. COHEN: Right, 3A and 3A1 are inverted, yes, thank you for that. So 3A should have a checkmark under two lags; 3A1 should have a checkmark under one lag. Thank you for that.

That's a good catch.

DR. DORAN: Let's all make that change to make sure we're all on the same page.

MS. FEILD: Shouldn't 3B have the covariate ELL, SWD, and attendance based on that?

DR. DORAN: Yes. Thank you.

DR. COHEN: Excellent. Thank you.

COMMITTEE MEMBER: Would you say that again, please?

DR. DORAN: Yeah, SWD under row 3B under covariates, in 3B under covariates, write SWD, ELL, and attendance. Under Model 3A in that row, remove the checkmark under one lag and instead put the checkmark under two lags. In the row below it, Model 3A1, put a checkmark in the column for one lag and remove the column for

American Court Reporting
850.421.0058

two lags. Thank you for that catch; I'm sorry for the error.

DR. COHEN: Okay, so there are four dimensions to the models. Differences versus covariate, four versus everybody else. Only teacher effects and school effects, Model 1 versus Model 3, the number of prior year achievement -- those are called lags -- and the particular covariates. So these are the four dimensions you want to be thinking about.

DR. DORAN: Yes, go ahead?

MS. FEILD: I have a general question.

When we talk about your data, are we only talking about students that have been promoted? In other words, if you're looking at three years of data, would that be progression from 3rd grade to 4th grade to 5th grade, or would we be including data 4th grader retained, 4th grade, 5th grade?

DR. DORAN: Christy, how did we do retained students? Did we have -- if we had, say, a 5th grade student who was retained, would we be using the retained data -- would we use two prior years of data or would we only use the data if they were promoted in this sequential

American Court Reporting
850.421.0058

order?

DR. HOVANETZ: We would use the prior years of data or would we only use the retained data -- would we use two?

DR. DORAN: Any scores.

MS. FEILD: Regardless of the grade level.

DR. DORAN: I don't remember that -- (inaudible) -- if we use any prior school --

DR. HOVANETZ: With the exception of 3rd graders, we do not use any 3rd graders.

MS. FEILD: What about retained 3rd graders? No?

DR. DORAN: No. Remember, we cannot estimate teacher effects in 3rd grade because there's no prior achievement data. So the only -- you have students who have two 3rd grade scores because they were retained for some reason, you'd get biased effects because of that.

MS. BROWN: On the chart 3C in the covariate, it lists gifted but I don't see gifted on the slide as a covariate.

DR. DORAN: Gifted is in the model. It should be on the slide.

DR. COHEN: It's actually in every model that includes SWD, it also includes gifted.

American Court Reporting
850.421.0058

DR. COHEN: It's actually in every model and 3A1 may be inverted, based on what's on the chart and it appears that the lags for 3A

MS. EDGECOMB: What will help the process plus and minuses. Do we each do these individually? Is that correct?

DR. DORAN: That's right.

MS. EDGECOMB: What will help the process so that we limit subjectivity in doing this?

DR. DORAN: Good question. When we talk about precision, we're going to show what we need by precision and we're going to show you results. We'll actually show you how the models vary in the terms of their precision. We'll find precision. Then we'll show which models are more precise and less precise, and then you'll be able to make your judgment based on data. Which of the models are more or less precise? Each of these criteria are associated

American Court Reporting
850.421.0058
with data, so there's no virtually no
subjectivity in that regard. Whether or not the
models are precise is an empirical question.
Whether or not they are school effects, we show
you the consequences of school effects or not,
and there's data. Same thing with parsimony,
classification, accuracy, and lags. There are
data that we will present associated with each
of these criteria and you can make a judgment on
whether Model 1 is better than Model 2 based on
the results of what we show you.
Christy?
DR. HOVANETZ: Just for clarification, this
is just an advance organizer for you all to take
into -- we're not necessarily going to ask you
to keep track of points or numbers, but just for
you to be able to reflect in an organized way
and that we're all doing it the same way. So
when we're talking about Model 3A, you can look
at 3A and see the notes that you've taken for
that specific model to help refresh your memory
because there are seven different models that
have not -- very fancy names.
MS. BROWN: I just want to clarify the
American Court Reporting
850.421.0058

I'm going to actually get a glance -- this
is a slide we talked about last time. I already
mentioned this when I was talking about the
differences between the fixed and random effects
earlier. But I want you to recall that fixed
and random effects are different ways of
estimating the same thing. In fact,
mathematically we know that they're expected to
be the same. They're expected to be no
different from each other under the condition
that classes get bigger. You have a lot of kids
in a class, then the numbers start to look about
the same. How many? I don't know exactly, 20
or 25 or 30 kids in a class; those numbers
should convert to the same value.
The reason we estimate these different
models is because there are different classes of
statisticians, some people who like to treat
these models as fixed effects, some who like to
estimate the most random effects. There are
statistical nuances, there are certain
properties of the random effects that people
like in certain properties of the fixed effects
that people like, and essentially what we've
shown is that when we remove those teachers who
American Court Reporting
850.421.0058
was a variable that was within the teacher’s evaluation, we talked about whether or not it
When we talked about including variables in the models to begin with; and just to refresh
our memory as to the discussions that we had on April 4th and 5th and again on the 14th at our
webinar, the reason we are looking at adding controlled variables is to reduce the variances'
unequal distribution of students that (inaudible) in teachers’ courses. There’s limited debate. We had this conversation back on the 4th and 5th and we did on the 14th about whether or not adding a lot of controlled
American Court Reporting 850.421.0058
variables is going to make a difference in the model, whether or not it’s going to make it more precise and whether or not it’s going to actually level the playing field for teachers.
Some of the rationales for including student characteristics is to eliminate that bias, but the policy implications of it is student who want to set it for differing expectations for different students. So just a little reminder or refresher about the conversation that we were having on these variables before.
Reminder on the framework that we operated under when we were talking about which variables to include. First of all, we went through and we looked at the variables that were in Senate Bill 736, the SWD, the ELL, and the attendance, and then we had the brainstorming session where you all listed out 20 or so variables that we had conversations about initially on the 4th and 5th and then went through each of them in detail on the 14th and made judgments about them then. When we talked about including variables in the evaluation, we talked about whether or not it was a variable that was within the teacher’s
classroom.  When we talked about variables being within the teacher’s classroom, we talked about whether or not it was a variable that was already being measured by another variable that we were looking at, and also whether or not it could be explained by pre-test data. That was the framework that we were operating under to put variables into the model for evaluation.
We evaluated a lot of variables where we thought let's just see what this looks like and we'll base our judgments on the results, but we will be using the same framework as we're considering the results of these variables, not just whether or not they're significant but whether or not it does make a difference in the precision of the model and whether or not policy-wise it's the right variable to be included. So just keep those same conversations in mind that we had before.
Okay. This is the list of variables that have been evaluated within the models. Students with disabilities status was done with a dichotomous variable for each of the individual disabilities. So we can either play a game where you guess which variables are D, E, Z, or I can just tell you. We did not include the exceptionality codes listed here based on the framework that we were operating under to put this.
codes, we had talked about looking at primary and other exceptionalities. This is only looking at primary student disability. We did not look at secondary or other exceptionalities. It's only primary.

For English language learner status, this is also a dichotomous variable. Students have to be coded as LY or currently receiving ESL services, and they can only be coded as LY for two or fewer years. So if a student is in his third year of receiving services, they are not considered an ELL student for our purposes.

Yes.

MS. ACOSTA: Just to clarify, the LY classification has really no bearing on their ESOL level, so in other words they could have been receiving for two years and have reached ESOL level four or still be in ESOL one, and they'll be treated equally?

DR. HOVANETZ: Yes.

For attendance, we treated attendance as a continuous variable the number of days the student was in attendance at the school. So if a student was in multiple schools, we added the number of days in attendance in all of those schools to get the total number of days the student was in attendance. It's a continuous variable, just the number of days present. Just as a reminder for this particular variable, this is information that comes in during survey five which is end of the summer and so for the school year, it's counted as a transition. We did not use state determined class sizes. We actually just counted the number of students enrolled in a course and that was the number of students or the class size for that particular course.

Homogeneity of class composition. Harold mentioned this already, but in order to determine the homogeneity of a class we looked at the score at the 25th percentile and the 75th percentile and took the difference. If it's a small difference between the 25th and 75th percentile, it's a homogenous or more homogenous class, a larger difference is a less homogenous class and this was done on a continuous variable. The mobility calculation, also a continuous variable, we looked at the number of transitions a student made from school to school. If we have only one record for the student, the student had zero transitions. If we have two records for the students in different schools, that's considered one transition. So each time a student changed schools during the school year, it's counted as a transition. We did encounter some students that had two records with two entry dates into the same school. If there was a 21-day period between the exit date of the school and the following entry date into that same school, they were considered to have made a transition. If they spent time somewhere else, it may not have been in Florida, but we're not at that school that actually made transitions.

Age is also a continuous variable. We looked at and calculated the modal age for the grade as of September 1st and took the difference between the modal age and the student's actual age to come up with the age difference.

And I will turn it back to Dr. Doran.

DR. DORAN: All right.

DR. HOVANETZ: Well, first, do you have any questions about variables?

MR. MOREHOUSE: I have a question about how you define homogeneity. Is it devised strictly on test scores?

DR. HOVANETZ: Yes, strictly on prior student achievement.

MR. MOREHOUSE: And what do you mean by "unique course" or "each unique course"?

DR. HOVANETZ: For a unique course, we looked at the district and the school, the courses offered, the course number, the period the course was offered, and the teacher for which the course was offered. So a unique course is any course number that is unique by school and by district and period.

Lance.

MR. TOMEI: I have a question about the class size statistic. How are the data collected and how stable is that particular statistic between data collection points?
DR. HOVANETZ: We are only looking at survey two data and survey three data, so it's that second week in October and that second week in February, and it's whichever students are enrolled during those two specific weeks in that particular course in that school, in that district with that period number. That's considered the number or the count of enrollment for class size.

MR. TOMEI: I'd like to ask the P-12 reps on the committee. Do you see that as a fairly stable statistic or is that one where there could be variance in the statistic itself that's not going to be captured in the reported data that will be used in the model?

MS. FEILD: Why not -- it was sort of a combination question. First, what was the self count within the class period teacher that was used to either aggregate the data or not? Secondly, how are we handling semester courses in the high schools? Generally, the kids are FTE's, did we account for that?

DR. HOVANETZ: Yes and no. So unfortunately we don't have a minimum class size to a discussion that we can have, but based on American Court Reporting 850.421.0058

DR. HOVANETZ: I have an entire course code directory that I can show you there are 166 courses that were polled for reading and 90 courses that were polled for math, and it's based on the Department's determination in the course code directory of what is a math course and what is listed as a reading or English language arts course. And in your packet it shows specifically there's a course code directory on how those determinations were made by the Department. It's a separate handout that was behind your Power Point on the right-hand side and the reading courses are courses that are identified as requiring a reading certification to teach them or a course that is mandated by the State Board of Education as a remedial course, a remedial reading course.

Math courses are identified by the prefix, those are a little bit more simplistic to identify. English language arts was determined by a committee and then also by the course code prefix. So I have a list of all the courses that were included for the particular analysis.

MS. YOUNG: I remember the discussion about attendance, but I didn't remember the outcome. American Court Reporting 850.421.0058

1 For students that are in multiple courses during the day and they have a tendency to leave after lunch, did we -- do we have data for that or just the whole day or they're marked present for the whole day, that's it?

DR. HOVANETZ: Yes. There was no way to capture by course at this time.

MR. COPA: Christy mentioned this but just one thing to add. As we talked last time in April, we have limitations on what we currently have to model, but one of the outcomes of one out of this process is to look at different ways to capture the data, improvements in the data systems, and we have already started conversations at the department level about capturing a lot of this information such as attendance at a course level as opposed to the daily attendance that's currently collected.

MS. FEILD: Referring back to Lance's question, I think there will be issues. Probably not monumental, but I know there are districts that offer some of the core courses within a semester. For example, algebra which is not going to be dealing with an end of course; a lot of districts are planning on American Court Reporting 850.421.0058
offering it as a semester so students can request --. So we'll have to take that into account.

DR. HOVANETZ: And that's okay, though, because we will still have the district, the school, and the course number for that particular student. So we do have that information right now and in the evaluation.

Other questions? Ask as many questions as you want today. Today is all about just getting you information, making sure that you're comfortable with the results based on the models we've selected or you selected, and we want to get all the information because we'll start making decisions tomorrow, or you'll all be starting to make decisions tomorrow about which ones we're leaning towards.

MS. EDGECOMB: In one of the earlier slides when you began talking this morning, you talked about the importance of data availability and accuracy. Is the assumption that all districts have in place the capacity to provide those two characteristics about data and input?

DR. HOVANETZ: That's a great question. I think a lot of the data that we are using is accurate because it's been used for other purposes. There are some pieces of information that will hopefully get more accurate as we continue to use them, but this is the best available information and as we select which variables to be included and highlight that with the districts that there will be more attention paid to insuring the accuracy of those. We'll also have quality check processes in place where districts are signing off on their data, just as they do with other school accountability measures for school grades and AYP, so they'll have the opportunity to review the data before these results come out. But it's a great question.

MS. EDGECOMB: Okay. Can I continue just a little bit with that? Is there going to be a timeline when all the districts do or perhaps for those districts who aren't up to par that will give them the opportunity to develop some systems that, data collection or dashboards or whatever, to get those things in place knowing that these are the expectations?

DR. HOVANETZ: Sure. As Juan has mentioned, they are in the process of revising...
Dr. Hovanetz: Yes, this is based on the courses the student are holding.

Ms. Feild: I could go on. I could ask another question. Elementary school, self-contained teachers who are teaching both reading and math would have two value-added scores, a reading score and a math score?

Dr. Hovanetz: Correct.

Ms. Feild: Assuming the school coded that as a self-contained?

Dr. Hovanetz: Correct.

Ms. Feild: So if in fact the school is departmentalizing it but they did not code the teacher as departmentalized, which we know they do, then the teacher will be attributed math scores in essence and in fact that teacher didn't really have math scores.

Dr. Hovanetz: Correct.

Ms. Feild: I think this alludes to Miss Doretha's comment on accurate scheduling and data pieces in the past; that was not important.

Dr. Hovanetz: Correct. We don't have actual runs of which teacher was assigned reading and which teacher was assigned math by a particular teacher, but one of the things we found is an elementary school teacher was more likely to have multiple teachers than a middle school or high school teacher was. Elementary schools are doing a good job of parsing out the students taking spelling, the students taking writing, the students taking reading, the students taking math; and so they have multiple course enrollments, it's not just fourth grade. So that's been a big shift in the data over the last five or ten years. But we are seeing that more students in elementary school have multiple courses associated with them.

Mr. Cop: Just another disclaimer again.

These are all great questions, and again back to the issue that we're operating under the constraints of what we have currently collected for this modeling purpose, but just so everyone is aware as well, both the law requires a roster verification process to be in place for this when this is operational, the Department with partner districts. Hillsborough is one, NEFEC is another one, and also Osceola County of developing a teacher-student data link roster verification system through a grant process with SELT, which is -- I don't know what -- I can't think of what the acronym is right now, but they have a grant through the Gates Foundation. So we're working with them over the next year and coming months and we'll be putting forth a process in place working with our district partners and open to the entire state on a roster verification system to improve that data that will be so fundamental to this purpose so that we can deal with those issues such as the variability in how schools or districts may report this course information, since it hasn't been used for this high stakes accountability purpose in the past.

Mr. Le Tellier: Going on with what you were saying with the count two and three if the student moves from one school to another, is the teacher that had them for the second count or the October -- would that affect their value-added model at the number three?

Dr. Hovanetz: That's a great question. That goes to the growth expectations and attributions, which is in a few slides, but maybe I'll skip to it now just to address your specific rule.

This is important for just kind of the fundamental understanding of how we determined expectations. Students' growth expectations are determined on the courses they are enrolled in, so if the student is enrolled in a reading course their expectation is based on that reading course. If a student is enrolled in multiple reading courses, their expectation is based on multiple course enrollments regardless of which teachers they had. It's based on the number of courses the student has taken in that subject to develop the student growth expectation. Harold will talk a little bit more about what those expectations are, but essentially we looked at students enrolled in 1 to 6 reading courses, 1 to 6 math courses and American Court Reporting 850.421.0058.
That is not very homogenous. Then I move to a homogeneity variable that had a large number. Skills, in the first school I would have one -- diverse in terms of kids entering math. A variable is calculated for each class. I'm taking Algebra 1 in both schools; that's one and move halfway through the year to school B.

The attribution of that growth is given to each teacher that student had. So if the student's growth expectation is based on one course and they've had one teacher, that teacher is fully accountable for that student's growth. If a student is enrolled in two or more courses, that growth expectation is a little bit higher and that teacher -- both of those teachers that student had -- whether it's the same teacher or different teachers, both of those teachers are accountable for that higher growth expectation.

So if Gisela and I both had reading courses -- she was teaching one reading course and I was teaching another and Stephanie was in our class, she has a higher growth expectation because she is taking two courses. I'm fully accountable for what you do in meeting that higher growth expectation.

American Court Reporting 850.421.0058

expectation and Gisela is fully accountable for you meeting that higher growth expectation. So we're both accountable for that higher expectation.

MS. FEILD: So how would the homogeneity of the class composition -- how does that work into a double portion? You'll have a lot of kids who'll have English and intensive reading. So how does that variable work?

DR. HOVANETZ: The homogeneity is a course level variable and it's generated for each particular course and it's used as a controlled variable as a model.

DR. COHEN: So we drove right into the fine details. So there's a difference between a course and a class. I'm a student in school A and move halfway through the year to school B, I'm taking Algebra 1 in both schools; that's one course. It's two classes. The homogeneity variable is calculated for each class separately. So if I was in a very broad diverse -- diverse in terms of kids entering math skills, in the first school I would have one homogeneity variable that had a large number. That is not very homogenous. Then I move to a course level model.

The only time it makes a difference is when you're teaching kids with a differential number of courses. So I have some of my kids who I'm their only teacher for, they count one; I have some kids who are in my class and in another reading class. Then do they count one or when I calculate my average do they count as one kid or do they count half as much as the other kids because I would only be half-teaching them? The decision that we understood from the last meeting was that all kids should count equally no matter how many other teachers are teaching them. That was the attribution that Christy was talking about.

Are there questions about that?

MR. LeTELLIER: I don't think I quite got what I was trying to say, my point, across.
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>course in that school. So those are the kinds</td>
<td>of recommendations we can make to the -- this</td>
</tr>
<tr>
<td>2</td>
<td>a course in that school. So those are the kinds</td>
<td>group can make to the Department. You all have</td>
</tr>
<tr>
<td>3</td>
<td>of recommendations we can make to the -- this</td>
<td>the ability to say we'd like to see it done this</td>
</tr>
<tr>
<td>4</td>
<td>group can make to the Department. You all have</td>
<td>way, but right now this is the best available</td>
</tr>
<tr>
<td>5</td>
<td>the ability to say we'd like to see it done this</td>
<td>information. Nicole?</td>
</tr>
<tr>
<td>6</td>
<td>way, but right now this is the best available</td>
<td>MS. MARSALA: Does it have to be the same?</td>
</tr>
<tr>
<td>7</td>
<td>information. Nicole?</td>
<td>I mean, do we have to say -- I mean, maybe for</td>
</tr>
<tr>
<td>8</td>
<td>MS. MARSALA: Does it have to be the same?</td>
<td>middle school and elementary where we don't have</td>
</tr>
<tr>
<td>9</td>
<td>I mean, do we have to say -- I mean, maybe for</td>
<td>the block scheduling; can it be two and three</td>
</tr>
<tr>
<td>10</td>
<td>middle school and elementary where we don't have</td>
<td>where they have the block scheduling it could be</td>
</tr>
<tr>
<td>11</td>
<td>the block scheduling; can it be two and three</td>
<td>two or three, or does it have to be the same</td>
</tr>
<tr>
<td>12</td>
<td>where they have the block scheduling it could be</td>
<td>everywhere?</td>
</tr>
<tr>
<td>13</td>
<td>two or three, or does it have to be the same</td>
<td>DR. HOVANETZ: Again, that's a decision</td>
</tr>
<tr>
<td>14</td>
<td>everywhere?</td>
<td>that the committee is going to be able to make.</td>
</tr>
<tr>
<td>15</td>
<td>DR. HOVANETZ: Again, that's a decision</td>
<td>You all can decide how that's done. If you want</td>
</tr>
<tr>
<td>16</td>
<td>that the committee is going to be able to make.</td>
<td>it to be that that student has to be with that</td>
</tr>
<tr>
<td>17</td>
<td>You all can decide how that's done. If you want</td>
<td>particular teacher for the entire year during</td>
</tr>
<tr>
<td>18</td>
<td>it to be that that student has to be with that</td>
<td>elementary school, again it's a decision</td>
</tr>
<tr>
<td>19</td>
<td>particular teacher for the entire year during</td>
<td>entirely up to the committee.</td>
</tr>
<tr>
<td>20</td>
<td>elementary school, again it's a decision</td>
<td>MR. LeTELLIER: Is this something that can</td>
</tr>
<tr>
<td>21</td>
<td>entirely up to the committee.</td>
<td>be made at some point as a business rule of some</td>
</tr>
<tr>
<td>22</td>
<td>MR. LeTELLIER: Is this something that can</td>
<td>sort? In talking with the Department, you know,</td>
</tr>
<tr>
<td>23</td>
<td>be made at some point as a business rule of some</td>
<td>whether or not this could be something</td>
</tr>
<tr>
<td>24</td>
<td>sort? In talking with the Department, you know,</td>
<td>implemented that we can look at because you were</td>
</tr>
<tr>
<td>25</td>
<td>whether or not this could be something</td>
<td>American Court Reporting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>850.421.0058</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January or February or whatever time period, and</td>
<td>that child is now going to count in the</td>
</tr>
<tr>
<td>2</td>
<td>that child is now going to count in the</td>
<td>calculation --</td>
</tr>
<tr>
<td>3</td>
<td>calculation --</td>
<td>DR. COHEN: That is a policy decision and</td>
</tr>
<tr>
<td>4</td>
<td>DR. COHEN: That is a policy decision and</td>
<td>I'm going to hand that over to Christy. I don't</td>
</tr>
<tr>
<td>5</td>
<td>I'm going to hand that over to Christy. I don't</td>
<td>know what data we have to support it, but that's</td>
</tr>
<tr>
<td>6</td>
<td>know what data we have to support it, but that's</td>
<td>the kind of recommendation I think the committee</td>
</tr>
<tr>
<td>7</td>
<td>the kind of recommendation I think the committee</td>
<td>can make.</td>
</tr>
<tr>
<td>8</td>
<td>can make.</td>
<td>DR. HOVANETZ: Yes, absolutely, and that</td>
</tr>
<tr>
<td>9</td>
<td>DR. HOVANETZ: Yes, absolutely, and that</td>
<td>goes back to Gisela's point, too, is if you do</td>
</tr>
<tr>
<td>10</td>
<td>goes back to Gisela's point, too, is if you do</td>
<td>require the full year attendance and it's a</td>
</tr>
<tr>
<td>11</td>
<td>require the full year attendance and it's a</td>
<td>semester long course, is that appropriate? And</td>
</tr>
<tr>
<td>12</td>
<td>semester long course, is that appropriate? And</td>
<td>so being able to look at how we're capturing</td>
</tr>
<tr>
<td>13</td>
<td>so being able to look at how we're capturing</td>
<td>attendance, data, and information, and being</td>
</tr>
<tr>
<td>14</td>
<td>attendance, data, and information, and being</td>
<td>able to make recommendations to the Department</td>
</tr>
<tr>
<td>15</td>
<td>able to make recommendations to the Department</td>
<td>on how to revise that is going to be essential</td>
</tr>
<tr>
<td>16</td>
<td>on how to revise that is going to be essential</td>
<td>here.</td>
</tr>
<tr>
<td>17</td>
<td>here.</td>
<td>Keep in mind as we're doing this, this is</td>
</tr>
<tr>
<td>18</td>
<td>Keep in mind as we're doing this, this is</td>
<td>with the data we have available and that was</td>
</tr>
<tr>
<td>19</td>
<td>with the data we have available and that was</td>
<td>what we talked about the last time we were</td>
</tr>
<tr>
<td>20</td>
<td>what we talked about the last time we were</td>
<td>together and this is the best information we</td>
</tr>
<tr>
<td>21</td>
<td>together and this is the best information we</td>
<td>have right now. You all are bringing up great</td>
</tr>
<tr>
<td>22</td>
<td>have right now. You all are bringing up great</td>
<td>points in how to capture additional information</td>
</tr>
<tr>
<td>23</td>
<td>points in how to capture additional information</td>
<td>by course, by period, you know, by number of</td>
</tr>
<tr>
<td>24</td>
<td>by course, by period, you know, by number of</td>
<td>days actually in attendance in that particular</td>
</tr>
<tr>
<td>25</td>
<td>days actually in attendance in that particular</td>
<td>American Court Reporting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>850.421.0058</td>
</tr>
</tbody>
</table>
that because I've heard a lot of feedback from
teachers who have watched the webinar and
language arts is taught quite differently when
you're focusing on writing.

DR. HOVANETZ: It's a fantastic point you
make. Tomorrow afternoon we have a conversation
about the course code directory that we don't
want to overwhelm you with now, as we're trying
to facilitate the process of having you all make
recommendations to the commissioner. But for
the process of this evaluation, we use the
information the Department had for -- course
code directly. The Department recognizes that
this new information needs to be revised or evaluated at least
to determine are these appropriate courses to
have for the -- evaluation of teachers on their
reading FCAT and on the math FCAT. And we have
a list, the master directory of all the courses
that were included. Like, I believe there are
166 for reading and 90 for mathematics that were
included for purposes of this evaluation.

This summer, this committee's
responsibility when we get together next will be
to talk about which courses should be the ones
included or required for the FCAT statewide
American Court Reporting
850.421.0058

evaluation. So they're bringing up fantastic
points. The Department has fully recognized
that's for tomorrow.

DR. COHEN: One thing that folks are
concerned, well, the kid's only in my class
for half the course, how can that be attributed
to me? We don't know if it's half the course
unless we know whether it's a full year or a
half-year course, right? Gathering this data
and - and the committee, I would recommend that
the committee think for a couple of minutes
about that issue and which data you need to
support the kinds of analysis, the kinds of
attribution you want because right now that's
not in the state data system. Particularly it's
not --

PANEL MEMBERS: (Over-speaking.)

DR. COHEN: So we could have.

DR. HOVANETZ: The only thing we don't know
with the identified data set is there is not a
term that says that this is a full year course
or if this is a semester course. So if a
student is repeating a high school course, it
will look to us as if it's a full year course
rather than repeating a semester course, but
that's an infrequent occurrence.

DR. COHEN: We know how long kids were
American Court Reporting
850.421.0058
DR. HOVANETZ: Correct.

MR. COHEN: So we don't know about block schedules. We don't have that.

DR. HOVANETZ: We don't technically have that information. We have what district and school and course number and period number and we have which survey the student was enrolled in; and if they're enrolled in the same one identical information for two in survey 2 and survey 3, the assumption is it's a whole year course. If they're only enrolled it in one semester in survey 2 or survey 3, the assumption can be it's a semester long course, but there's no actual data that says semester or full year.

MS. BROWN: I'm confused and this will ask everybody else. At least in our scheduling system that we report, there's a cell for term -- term 1, term 2, or term 3. Term 1 is for semester, term 2 is second semester, term 3 is a year long course. So I don't understand why that data element isn't available to you.

DR. HOVANETZ: We'll get that data --

MS. BROWN: I think part of the problem is sometimes schools semester-ize courses for GPA, for kids in graduation; and if that happens after the collection period it's possible there could be some error in that, but you're right.

MS. FEILD: But in general you should be able to know. You should clearly be able to see how that student is scheduled in that reading course.

DR. HOVANETZ: Okay. We will be sure to --

MR. MOREHOUSE: I have a question.

Hillsborough County, for a term 3 full year course, students get a grade for the first semester, second semester; and those grades are based upon what?

MS. BROWN: What do you mean by --

MR. MOREHOUSE: Is that material that's on that first semester?

MS. BROWN: Correct.

MR. MOREHOUSE: Then they evaluated on that. So why didn't -- wouldn't that be a test score that's scored?

MS. BROWN: Because FCAT scores are only given once a year, so that's the reason that you only have that one option. But, I mean, that's where we're trying to attempt to get at attribution more on a semester level instead of

DOC in December. So it's really -- sometimes people are we talking about? Is this very common that you only have a semester course that's going to be tested on the FCAT?

MS. FEILD: I think part of the case in Miami-Dade is, of course, recovery. They're even talking now in terms of algebra making it a semester course for course recovery, taking the

MS. ACOSTA: I have a question actually, maybe some of you can address this as well. I think you partly answered it, Anna. I was curious how many kids are actually impacted by having semester courses for FCAT, for things to get tested on the FCAT because at our school it would be a rare student who would be enrolled in a year-long course that had an FCAT test. Do you see what I'm saying?

How many students are we talking about? Is this very common that you only have a semester course that's going to be tested on the FCAT?

MS. FEILD: I think part of the case in Miami-Dade is, of course, recovery. They're even talking now in terms of algebra making it a semester course for course recovery, taking the

kids pass the first semester but fail the second semester, so they're course recovery-ing at the beginning of the next year just the second semester or vice versa. So there's thousands in Miami-Dade, thousands who fall in that criteria.

MS. BROWN: And what we see is that is definitely the situation in our district, and there are many, many students; but when you also consider mobility which I know we're capturing, but when students transfer school to school even in elementary, when you're highly mobile you show up in multiple places but it's very important because you're still one FCAT score, and so which teachers are given attribution for that effort? It's a lot of students.

MR. TOMEI: I just want to make an observation; I think this is an extremely important conversation that we're having right now for a couple of reasons. First of all, over the course of this committee we're going to get past the point where we're only talking about FCAT scores as a measure, so Lawrence's comment is an important one to keep in mind. What I'm very concerned about is that the ultimate model that gets put in place is going to be used for
1. multiple purposes. It will be used to evaluate
teacher effectiveness; it will be used to
evaluate school and possibly district level
effectiveness; and it will be used to evaluate
teacher preparation program efficacy. What may
work at some levels -- and I will go back to the
class size -- the question I asked about class
size and how stable that is between the
measures. When you get a large enough end, if
you're using data to evaluate a teacher
preparation program that puts out a thousand
students a year, it's probably not an issue.
But for one individual teacher it could be a
really big deal if that statistic is unstable.
So we've got to keep in mind to protect the
transparency and the integrity of the model for
individual teacher accountability. I think
that's the lowest common denominator and I think
that has to be an important outcome of this
committee.
So these conversations and comments like
things are infrequent, well, if you're that one
teacher that experienced that infrequent event,
that doesn't mean a lot to you. We need to
protect that teacher as well.  
American Court Reporting
850.421.0058

2. So again, I just want to emphasize that I
think this is an extremely important
conversation we're having.

MR. LeTELLIER: That goes back to the last
time we met the reason why we took the norm
referenced model and threw it out the window,
the quantitative -- I think it was called the
quantitative based because we were concerned
about that lower two percent of teachers that no
matter how good they were doing, they were in
that lowest two percent, so you don't have a
chance of getting support as effective, and I
think that that is something that I will agree
with that that we can't take the risk of just
saying, well, we may have three or four
teachers, I'll just put out a handful, that they
may go by the wayside because that's certainly
not fair to the teacher and it doesn't speak to
the integrity of how we evaluate people.
So I agree with that and I think that the
end result that I would see is moving towards a
system that would go with the individual teacher
for what they are doing so they're accountable
for their work, and I think -- you know, in
Florida we're all talking about aligning
American Court Reporting
850.421.0058

1. everything to business models. You know,
accountability is kind of something we haven't
talked about too much, but all this stems out of
a business model of being accountable for what
your job is and making sure that you're reaching
those goals or exceeding them. Well, if we're
truly going to do that, it would have to be
something that that teacher is able to reach.
So I just wanted to piggyback off of that.

MR. Tomei: I think part of that overall
goal has to be that we need to do everything we
can within the legal limits that we have in
terms of what variables we can put in the model
to mitigate to the greatest extent possible the
unintended consequences of discouraging the best
teachers from going into the locations that are
in most need.

MR. LeTELLIER: Absolutely, and I just
thought of this about an hour ago, is we had
talked about unintended consequences. I can't
remember what you -- what's the word you used
last time?

PANEL MEMBER: (Inaudible.)

MR. LeTELLIER: And I've been dying to say
this and I haven't said this publicly, but I
American Court Reporting
850.421.0058

1. think we should: Every profession is going to
have people that try to play the game, that are
going to try and skirt around things and try and
make it work for their benefit, but I would like
to think that the majority and not just a simple
majority, but the vast majority of teachers are
not going to try to game the system. So what I
guess we don't want is the reverse where we make
it so that, okay, we've taken away the perverse
incentive but we've made it so that it's not
sustainable or obtainable to the teacher that is
trying to do their best. I think that's
something that's been weighing on my mind for
the past month is making sure that we don't just
rule out things so we say, well, there's a
pervasive incentive there for a handful of people
that are going to take advantage of that and
maybe there's some way that we can address that
rather than just throwing out a specific
variable that we figure, okay, how can we
address if that happens.

But I really -- you know, being a teacher
for the years that I have been and the people in
this room, the teacher, the superintendents, I
mean, the parents here, you guys know the
American Court Reporting
850.421.0058
<table>
<thead>
<tr>
<th>Page 78</th>
<th>Page 79</th>
<th>Page 80</th>
</tr>
</thead>
</table>
| 1 majority of teachers care, we love what we do.  
2 We want to help kids and the last thing on our  
3 minds is how do we game the system. So I think  
4 that's very important, something that I just  
5 want to share.  
6 MS. EDGECOMB: In this discussion and maybe  
7 I'm not hearing it correctly and maybe I  
8 shouldn't get bent out of shape about it. I'm  
9 worried about uniformity I'm hearing. Is that  
10 something -- because I hear how people are  
11 capturing and coding and talking about courses  
12 and when they end, how they end, and what they  
13 are and what they are not. If we're using the  
14 motto that's supposed to be fair and formalized  
15 instruction to capture and save information,  
16 isn't there some importance and value in some  
17 uniformity here?  
18 DR. HOVANETZ: There is. I absolutely  
19 agree. The Department has a directory that  
20 defines each of the particular variables that we  
21 are using, so I think they are well defined.  
22 It's assuring and again they have been reporting  
23 these variables to the Department for in many  
24 cases decades. In the same way with actual data  
25 elements, definitions, and particulars on how  |
| 80 | 1 technical assistance with MIS directors and  
2 others who are reporting on that information up  
3 to the state to be sure that they are following  
4 what's in the data element dictionaries.  
5 MS. BROWN: If I can just piggyback on that  
6 because uniformity is so important to the  
7 validity of the entire process, and in my  
8 experience what I can see the writing on the  
9 wall where we're going to need to go is how we  
10 clearly define those business rules for  
11 eligibility because in some cases we may not  
12 have uniformity in say, for example, course  
13 assignment. But we can create uniformity by  
14 creating the right business rule that looks at  
15 those in an either/or situation, but then also  
16 the process itself starts to find the anomaly  
17 and when we start to find the thing that stands  
18 out as different then there's just that  
19 oversight that starts to correct everything  
20 towards uniformity.  
21 We've discovered that in our own district  
22 when we're looking across large numbers of  
23 schools and how individuals are scheduled, et  
24 cetera, and we would do that major oversight and  
25 you think that you have every one common, and  |
| 81 | 1 they're supposed to report it and when they're  
2 supposed to report it, each one of them is very  
3 defined.  
4 It's assuring that they're following those  
5 proper procedures, which I'm assuming the  
6 majority of the districts are but continuing to  
7 reiterate as soon as we do determine which are  
8 the control variables or the covariates we're  
9 going to be including in these models to  
10 highlight those when they do that -- and I ask  
11 for -- when they're at the consortium meetings  
12 to be able to highlight and say these are the  
13 particular covariates that are included in the  
14 models, these are the data definitions that you  
15 need to be paying attention to, and then also  
16 through the review process to pay particular  
17 attention to these covariates, but I agree,  
18 uniformity is going to be paramount for insuring  
19 the accuracy of these calculations. I think  
20 that there are already a lot of processes that  
21 the Department has in place to insure that and  
22 continue to use this data and information that's  
23 going to improve the quality of it, too. But I  
24 absolutely hear that that is a concern and I  
25 think it's going to just take additional PD and  |
interpreting every district then we start
1 doing everything differently, and that's been
2 one of the problems across the state for years.
3 So transparency, teachers are asking -- when in
4 the summer? Are you going to do this late
5 summer so when they come back and they're in a
6 rush, won't even see the transparency?
7      DR. HOVANETZ:  Juan is panicking because of
8 what I said, but there's --
9      MR. COPA:  I just heard an audible groan
10 with the word "posted", but this is very --
11 districts have been very interested, of course,
12 as we're developing part of Race to the Top
13 their evaluation systems, information needed to
14 inform those decisions regarding their
15 evaluation systems, and we are committed to
16 providing districts with data on the model that
17 is eventually selected to help them form
18 decisions. So it's really a provision of data
19 to districts in a useable form to help folks
20 start to understand what it means, make informed
21 decisions about how to apply an evaluation
22 system. So that's the key point we'll
23 accomplish this summer.
24      MS. TOVINE:  Once those results are
25 American Court Reporting
850.421.0058

actually run and they go back, some of them are
1 selected, will this committee come back
2 together, say, in the fall or whatever time
3 period that is to kind of see how that's going
4 and maybe look for some of those problematic
5 areas and maybe could make some sort of
6 recommendation how to do it differently?
7      MR. COPA:  Gina, you didn't realize this
8 was a four year commitment? Very clear, this is
9 definitely not the last meeting of this group.
10 This group will continue to meet throughout the
11 four year process, not only dealing with -- we
12 have this -- there's the sense that this is the
13 end for the FCAT model as Cathy said at the
14 beginning; no, this is really the beginning and
15 so we will continue to review that. We will
16 start working on those other models, the EOCs
17 that will eventually come online starting with
18 Algebra 1 which was just administered last week,
19 so yeah, there's a lot of work to be done over
20 the next three years.
21      MR. LetTELLIER:  Juan, I think if I
22 understand you correctly and what I think will
23 make everybody feel a little better is what
24 we're looking at is presenting the data with the
25 American Court Reporting
850.421.0058

model that we use so that people can just look
1 and say, okay, here's where we're at. We tweak
2 it from there, nobody's jobs on the line because
3 of the data coming out in the summer. I think
4 that's what people are kind of --
5      MR. COPA:  Absolutely, yeah. This is for
6 information purposes.
7      DR. DORAN:  That's a good point. This
8 gives us the basis by which we can look at some
9 of the data, changing some of those business
10 rules wouldn't switch, for example, which
11 value-added model necessarily look different,
12 but it would change some of the things about
13 attribution. So it's still us; we're still on
14 safe ground on how we discuss what we intend to,
15 even when the business rules perhaps don't
16 change.
17      MS. FEILD:  Which FCAT scores are you
18 using?  The old scale, the ten year scale?
19      PANEL MEMBERS:  Over speaking.
20      DR. COPA:  Correct, for the purposes of
21 this summer you're well aware it's equated
22 exactly to the old scale, but it's part of their
23 work as well. It's no secret we're moving to
24 new standards beginning this fall. It will be
25 American Court Reporting
850.421.0058

applied starting with the spring 2012
1 assessment, so any change in the assessment will
2 likely result in further refinement of the
3 models as well. So again, on to this continued
4 responsibility of this committee.
5      DR. DORAN:  The models that we've
6 implemented are scaled independent, so even
7 should the scale change the models can still be
8 applied.
9      MS. FEILD:  No, I understand that but we're
10 making decisions on the model based on the old
11 FCAT, and it's possible that with the new
12 standard and the new scores the model we pick
13 now may not be the best model with the new data.
14 That's why I was asking and I think you answered
15 it by saying that we will be able to tweak the
16 model. What I suggest is that when we have the
17 new FCAT data and we almost have to wait two
18 years to have two years by this hearing, we have
19 to revisit the model because maybe at that point
20 some of the variables we're using are not
21 appropriate or we have to change the model.
22      MR. COPA:  Possibly, yes.
23      DR. HOVANETZ:  Absolutely. Four year
24 commitment.
25 American Court Reporting
850.421.0058
MR. COPA: These are all empirical questions. The commissioner must make a selection by June 1st. When that becomes part of State Board rules, I mean, that is something that would happen in the future going forward after we've had the work of this committee to look at years of performance, refinements, and models; and even when it becomes part of the State Board rule you could further refine that going forward, as well. So that's a key thing to keep in mind. It's not something that we will be etched in stone that cannot be changed. The anticipation here, of course, is this will be an evolving product.

MR. MOREHOUSE: I'm just curious. Is it possible to develop a description, a course description? There seem to be so many variations in terms of how a course is delivered, some courses will be graded one semester, some are year long courses, some are team taught, some are not taught; is it possible to have a description of those options for the committee so that any rules that we have to make in my mind, it's hard to depend upon how those courses are delivered and how they're graded and American Court Reporting 850.421.0058

when they're delivered.

DR. HOVANETZ: Yes, and I think we can fold this into our course code discussion that we'll have tomorrow, too, as part of the ongoing work of the committee as well. There are well defined course descriptions that have standards associated with them as well; the delivery method of those is more the district discretion on how the content is delivered, but we will definitely role that into part of the conversation. Doretha?

MS. EDGECOMB: I apologize if this is a stupid question up front, so if it's stupid I apologize.

Is there any value, and I think I heard Gisela say this, to recommend a model that had limited number -- this is progressive -- limited number of covariates and add onto it and over time have a lot of them and take away from those covariates?

DR. DORAN: I'll answer the question. There is a criteria that we will evaluate today that we call parsimony. Parsimony looks at whether not -- American Court Reporting 850.421.0058

20 minutes till.

(Whereupon, a short recess was had.)

DR. DORAN: The part of the conversation that we're about to go in right now is very data centric. We're going to start looking at some of the data relatively soon. We've got a couple slides to get through. The conversation we had this morning really should underscore how important the quality of the input data is into the statistical model. As I walked outside to go to the Einstein, which was closed to get a little snack, I looked at the guy laying the bricks and he's being so very careful to make sure that each and every brick fits right into place, and if it doesn't he pulls out, he's scraping the bricks, making sure they fit into place because in the end that thing has to look perfect. The reliability and the validity of these models rests tremendously on the inputs that go into this. Now that was an extremely important conversation and we'll continue that conversation; at this point we're going to transition a little bit in from some of those rules and decisions that were made at the last meeting, which were fully amended by the way. American Court Reporting 850.421.0058
There's no reason you have to hang your hat on
the things that were cited, as far as I know, in
terms of going forward when this becomes
operational.

But we're going to look now at some of the
data, and we will evaluate some of the models
based on the different categories that we've
laid out for you. So let's talk about a couple
of things and a couple of terms.

One, let's talk about this thing called a
deviation from an expectation. Given prior
scores and other characteristics of kids,
whether or not they're ELL, special ed, gifted,
so on and so forth, enrolled in two courses, we
have what is the average score of similar
students. That's what's called an expectation.

If you recall back to that scatter plot where we
had those regression lines, the expectation is
that line. Remember that line changes according
to students. So with students you have all
things being equal similar prior scores would
have the same expectation. Students who score
above that or below that have a deviation from
that expectation. In statistical terms that's
referred to a residual. It's the deviation from
American Court Reporting
850.421.0058

What score did the student actually get in
testing that's sometimes called an observed
score. It's the score that the student actually
received. So we have an expectation, we have an
observed score, and this is the deviation that
gets aggregated, not residual. Essentially,
it's aggregated and it's a little bit more than
a mean but it's equivalent to a mean. We take
those residuals and we aggregate them within a
teacher's class. We formed some decisions about
whether the teacher had high value added or low
value added.

Expected growth is a little deviation from
the expectation. Later on in the presentation,
we'll show some slides about differences in
expected growth on different sub-groups of
students, so let me define what expected growth
is. We have the expected score, like I talked
about here, which is the predicted score --
given your prior score and any other
characteristics about you, what is your
predicted score -- minus your prior score, the
first lag and the most recent lag. We call that
for each kid the expected growth. We have a
American Court Reporting
850.421.0058

predicted score and we have a prior score, if
you just subtract the prior score from that
predicted score for each kid you get an expected
growth. And that's a statistic that we formed
some summary statistics on and we will show you
later on in the presentation.

Christy already went over those.

Let me talk briefly about this. The
expected scores change in terms of their
definition depending on which model we look at.
Now, for each model we form an expectation, and
I'm going to put an extra word in there. We
form what is called a conditional expectation.
It's conditional on what your prior score was.
Kids with two different prior scores have
different expectations, but kids who have the
same prior score have the same expectation.
It's also conditional on whether or not you were
SWD, special ed or gifted or so forth. So, for
example, I think Gisela was asking this question
earlier, two kids who are exactly identical on
all of their characteristics -- prior score and
categorization in gifted, special ed, and so
forth -- they're the same on all of those. They
have the exact same predicted score, everything
American Court Reporting
850.421.0058

being equal, but you change one of those things.
So suppose students are exactly the same in
terms of their gifted coding and SWD coding and
everything, but they have two different prior
scores, they would have different expectations
and whether or not one student had a higher or
lower and depending on what their prior score
was. So all things being equal, kids with
similar prior scores have similar predicted
scores and similar growth expectations, but when
those things change they have different
expectations.

DR. COHEN: Harold, the numbering on this
chart is wrong.

DR. DORAN: In fact, you know what --
DR. COHEN: We have talked about this; this
is going to cause confusion.

DR. DORAN: Why don't I skip this because
this really doesn't add substance to some of the
criteria we're going to go through. We could
talk about those things contextually as we
encounter those models. Thanks, Jon.

All right, let's provide a big picture
before we start delving into some of the
specifics. Here's a big picture. There are a
American Court Reporting
850.421.0058
lot of models and many different variables. We have eight different models, seven grades in two subjects. There are more than 100 models that were estimated, and we had 18 different criteria by which we evaluate the models. If we wanted to spend time looking at each model for each grade for each subject against each criterion, we would need to be here for four years. This wouldn't be a four year contract; we'd be here for a very long time. It's not viable. Just simply cannot do it.

Now we're going to try and consolidate this information based on some things that seemed reasonable to do; I'm going to show you how we try to narrow some -- how we're facilitating this conversation today. The point here is that there is a lot to look at there are many different variants that were estimated from many different grades and subjects. But as we're going to show you, the key results are consistence across all grades. If we saw that models behaved differently across grades and across subjects, we would need to have pulled those models out and examined them. Why does grade seven Model 1 look so different than Model American Court Reporting

DR. COHEN: Yeah, I'll talk briefly about it. There will be a technical report that has all this in here, but we're going to -- when you estimate a model you want to make sure that the model is unbiased, right? We want to have unbiased estimates of teacher effects and we want to make sure that the way we're doing things statistically isn't introducing a bias.

What is a bias? If there's a true value then any deviation from that true value is a bias, right?

So in the real world we don't know how much value each teacher has. You can't know that about teachers. The teacher fairy can't land on your shoulder and whisper in your ear "Miss Jones is absent, add 10 points of value". But what we can do is we can make up data, right, and we can make up the data in different ways. So we know the true values. So what we do, we define a process for making up the data and then process the random ones. You choose random trials from this population. We generate 200 different data sets and then estimate the model 200 different times. Again, this is on data we have made up so we know what the true values are.

Is there a pen here? I'm always happier when I have -- look between, oh, yeah. Let me roll this out here. Okay.

So every teacher effect that we're going to estimate under any model, this is a value-added scale, and this is -- you think of it sort of as a probability. It's a density function of the
value-added, we'll just give it a little Greek symbol there for the value-added just to have something to put up there.

All right. So let's say a teacher's score is right in the center. So your value-added score is -- I don't know what the number is, maybe you're adding 20 points. That 20 points is a point estimate, but it's not really a point estimate because we don't know. All we have is we a probability distribution and they tend to come out to be normal probability distributions -- there are proofs that statisticians like to think about. So here's your point estimate.

Each point estimate comes with a standard error, right? So you know that the teacher is very likely to be at this point, a little less likely to be at this point, way less likely to be down there, really unlikely to be up here. But it's a probability distribution; it's possible.

So we estimate these standard errors. When we talk about precision, we're going to be leaning on these standard errors. If the standard error is smaller, the model is more precise. I have less uncertainty. If I had a curve like this, I would know with some certainty that this teacher was in at least in this range and did not fall out here. Is that clear?

So more precise is better. It sounds better, it is better.

So you can't observe the standard error, so there are statistical formula that you use and for these models they become very complicated.

So we tested the statistical formulas for the standard error by generating lots of data where we knew the real answer, and we tested -- we got the standard errors and we counted the number of times that the estimates fell out in this range. We counted the number of times the estimates fell out in that range and the same on the other side, because the standard error and the normal distribution tell us we ought to have five percent following here plus here. We ought to have ten percent following here plus here. For all the models that we're showing you, we wound up with unbiased standard errors. So the standard errors did -- when you make up the data and you look at it over and over, it does reflect -- the standard errors do capture and tell you the right percentage of teachers who:

---

DR. DORAN: It's easy. This standard error curve is more narrow, it's more precise. This means the estimate is more precise. If it's wider, it's less precise.

MS. BROWN: Did you mention something about accuracy, though? I thought -

DR. COHEN: All the estimates are giving us accurate estimates, unbiased. So we're taking accuracy and we're breaking it into two pieces, unbiased and precise. Is it centered around the right number and how much variation is there?

DR. DORAN: We're going to talk a lot about precision in just a minute here. So here's the roadmap at the bottom of the slide. There's actually a little bit more. This refers back to the crib sheet that we've given you. We're going to look at the model's different size, the effects attributed to teachers. We'll talk about that, how precise the estimates in terms of what they yield, what are the expectations of growth established for different groups of kids, and what is the impact of various models on different groups of teachers?

So this is big picture. Now we're going to start looking at some of the models across all different groups of teachers?
1 grades in both subjects, and this is going to be
2 the first thing that we're going to show you is
3 a line graph that shows the magnitude of the
4 effects. Let me actually just show you what
5 we've got here. We're going to start with
6 reading. This is all models, all grades. This
7 slide and the next slide are the only time today
8 that we will show you a statistic on all models
9 for all grades as we go forward and as I'll
10 explain in this particular slide, here this
11 slide provides the justification for why we only
12 look at one particular grade but all models. We
13 always look at all models, but only for a given
14 grade. In this particular slide, we look at all
15 models for all grades, okay. This is the last
16 time we'll do this, the only time we'll do it.
17 All right. Here on the X access we have
18 each grade and on the Y access we have what
19 we're calling the size of the effects. Now this
20 here, the model, remember looking at your sheet
21 differ in terms of a few characteristics. Some
22 of them only have teacher effects and some of
23 them have both the teacher and school effects.
24 We'll actually define what that means to have
25 both teacher and school effects in just a little
American Court Reporting
850.421.0058

1 bit.
2 But what we see here is that these are the
3 covariate adjustment models and they behave very
4 similarly. The effects are comparable across
5 all grades. We don't see the lines
6 criss-crossing in very unpredictable strange
7 ways. We don't see something that looks
8 extremely anomalous suggesting that Model 1 in
9 grade four is behaving very differently than
10 Model 1 in a particular grade. There are very
11 tiny differences in the models across grades.
12 So, for example, in this particular model here
13 which is, I think, on this square Model 1 it's
14 behaving similarly across all of these grades.
15 The effects always appear to be the largest.
16 We'll see why in a little bit. Then in here you
17 don't see criss-crossing.
18 Up here is the simple differences model.
19 Now it's measuring something different as we
20 explained before, all right? It's not measuring
21 the exact same thing as the other covariate
22 adjustment models are, and so we see a big
23 difference here between its effect and its
24 relationship with the other models, all right?
25 What we've done over here on the right,
American Court Reporting
850.421.0058

1 we're going to introduce this now but we're
2 going to look at this in a very detailed way in
3 just a little bit is what happens when we add
4 the school effects back into the teacher
5 effects. In part, we're not ready to evaluate
6 this criteria yet because we really haven't
7 defined why you should care about school and/or
8 teacher effects just yet. This is a high level
9 overview.
10 What happens if you add those school
11 effects back in; essentially what we see is that
12 the models behave similarly again. So adding
13 school effects back in as opposed to having only
14 teacher effects in some of the models causes for
15 the behavior of the models to be similar. Let
16 me go to the next slide because we see something
17 similar in math, and I'm going to revisit the
18 key point that we're looking for here. The key
19 point, the key take away in the slide that I
20 showed you before, in the slide that I'm showing
21 you now is that the behavior of the models
22 across grades for both reading and math is
23 comparable. They're similar. So we're using
24 this as a justification for why later we're only
25 going to look at a single grade all models.
American Court Reporting
850.421.0058

1 You're not ready yet to evaluate any of the
2 models based on the criteria, the precision, the
3 accuracy, and so forth just yet. This is just
4 the high level overview.
5 We see the same thing in math. We don't
6 see the model here and here and here and
7 criss-crossing lines which would suggest that
8 Model 1 behaves very differently depending on
9 what grade you're looking at. That's not what
10 we see. We see a consistent pattern here and
11 when we add the school effects back into the
12 teacher effects. It's the only inference that
13 we really want you to draw from these slides.
14 That is, models behave similarly across grades.
15 MR. LeTELLIER: Can I ask a question? How
16 large of a sample group do you look at because I
17 don't think we have that anywhere in the info?
18 DR. DORAN: It was -- I'll tell you, I
19 don't know the exact numbers but we do this by
20 grades. The number of teachers in each grade
21 ranges from 7,000 to 10,000 or so teachers per
22 grade --
23 DR. COHEN: It's all for the entire state.
24 DR. DORAN: Yeah, the entire state.
25 There's no sample, it's a population of kids.
MR. LeTELLIER: And the second thing is when we're looking at the effects of the variables when deciding this, I know now we're looking at sort of a macrocosm sort of dealing with the big picture; are we going to look at how those variables may affect an individual class?

DR. DORAN: Yes. Not a particular teacher, but on average across many teachers, yes. We see exactly that.

MR. LeTELLIER: Versus just plugging in to the whole state is what I'm saying because if you -- in other words, you may have with the attendance issue of a kid that missed 60 days out of the 180 days, that certainly hopefully is not the norm, but that may have great weight for that teacher in that class. So if we only plug that into those 10,000 teachers, it's going to be a blip on the screen and won't show up. That's my question.

DR. DORAN: All right. We don't look at any one teacher where there's any particular impact, so does adding that variable in change your ranking from high or low. But one of the things that we do look at is how correlated, American Court Reporting 850.421.0058

what's the relationship of the teacher effects across the different models. That tells us whether there's a flip-flop of teachers being highly classified and under one model then maybe being classified differently under another model. We do look at that. I can tell you now even though that's something we'll talk about a little bit later on, that the consistency of the classifications across models is virtually perfectly correlated. Okay?

MR. COHEN: Harold, we can say something -- when we get into covariates we can talk a little bit about what the likely effect of differences in rates of attendance would have on individual teachers' value-added lists. We can look at that. We can talk about it. We didn't pull out any individual teacher. There's data as to that question and again we're talking about covariates.

MR. LeTELLIER: I wrote it down. I get the general thing, but I don't know if anybody else is understanding where I'm coming from is that not missing something because it's in such a large set of data and making sure that it -- in other words, we make sure that a variable can affect a teacher in a certain way and we're going to take that into account. Now obviously if it only affects one teacher across the state, is that the norm? No. But what we're saying is if all teachers that would have a kid that missed 60 days, let's say, whatever teacher that was, is that an effect? Are we looking at that in that way?

DR. DORAN: Yes. Before we go into the different effects including attendance has on the predictions of the --

MR. LeTELLIER: Attendance was just an example --

DR. DORAN: Of the different covariates.

MR. LeTELLIER: Okay. Thank you.

DR. COHEN: I think it's math per grade says a kid who missed 60 days, everything else being equal, would have an expected scale score of seven points less. So you would have one kid who has an expectation of seven points less among your whole class. So it may matter and attendance is something that you want to think about. You can't learn if you're not in school, right?

DR. DORAN: One thing that I didn't tell American Court Reporting 850.421.0058

you at the beginning of the day is for grade seven. We brought data with us, some data files, so if there are things that you're curious about that we don't present or there are some questions that we can look at, if I'm talking Jon will crank through and maybe data analysis done or if somebody else is talking, I can crank through an analysis and we can try and answer some of those questions. There are no instances where we look at one teacher and look at whether or not that particular teacher is changing. We look at the population and how things are behaving across the state.

MS. WESTPHAL: When you're looking at the models, and I understand the purpose of this was to justify why we just picked seventh grade to look at, I see it in the reading and I see it on that one. Is that not a significant difference on the left between sixth grade -- fourth, fifth, sixth grade versus seventh, eighth, ninth, tenth?

DR. DORAN: So there's an interesting finding here, all right. If I'm following what you're saying, the significant difference is that we see this downward trend.
MS. WESTPHAL: Well, no, the consistency of the models seven, eight, nine, ten versus four, five, and six.

DR. COHEN: No, I think the important thing to take away from this is that the model -- even though it's developmental scales, the scales are kind of different between grades, and so you don't necessarily expect the same numbers in each grade. What's important is the relative ranking of the models. Model 1 always sees the biggest teacher effect, Model 2 always sees the next biggest effect, so the lines are parallel. That's what you're looking for in this. There are differences in the estimated effect size across grades and there's a difference in the pattern between math and reading. But that is confounded with the differences in the measurement itself. We have run everything and we've looked at all the data and the big findings stay the same. We just didn't want to try and present, you know, two subjects times seven grades for this many models. You'd never be able to look at the data.

MS. WESTPHAL: Well, let's say we're looking at grade six. Two of the models or seven grades for this many models. You'd never be able to look at the data.

DR. DORAN: So let's actually start looking at some of these results and some of the data. We're going to structure the criteria around the following four issues. We're going to always at some of these results and some of the data. We're going to structure the criteria around the following four issues. We're going to always be able to look at the data.

MS. WESTPHAL: That makes sense.

American Court Reporting
850.421.0058

or something like that?

MS. BROWN: Or if we are forced to make one choice, we have to make a decision; so that's important.

MR. COPA: You're not constrained in that regard. If one model fits reading better than math, you can have a separate model for math, separate for reading.

DR. DORAN: So let's actually start looking at some of these results and some of the data. We're going to structure the criteria around the following four issues. We're going to always have a question, what are we looking for? What do we want to know when we talk about precision? Then we're going to give you a statistic. What statistic gives us some evidence. Then the third thing is what are we talking about in that statistic? How do we know? What are we using to judge it by? Then last why do you care?

Just sort of a simple this is why it matters to you.

Then we're going to show you data and then we're going to try to summarize what we observe in the data.

We're going to first talk about precision.

American Court Reporting
850.421.0058
Precision in terms of its questions, what characteristics of value-added models lead to more precise estimates of the teacher effects? The statistic that we're going to look at is called the standard error. We're going to look at the standard error of the teacher effects. Let me talk about what a standard error is. It's actually something that's very familiar in polling, so for example if you look at the president's popularity rating and the president's popularity rating is 50% plus or minus 3 percentage points. That's typically what we see, right? That means, you know, it could be -- it's a little bit more than this but it's somewhere between 47% or somewhere between 53%. We're pretty sure it's within a small range. That's standard error.

Now supposed the president's popularity is 50% plus or minus 20 points. Well, is the popularity when I say 20 points, 30, or 70? That's a big range. It's a big range of uncertainty. The standard error tells you I've got a statistic, the president's popularity is 'x' 50%, but how certain am I that it's within that range. It's a certainty statistic. Well, American Court Reporting 850.421.0058

If the standard error is big, we don't know. If there's a big range, it could be anywhere in there or is there a small range? I actually know it's in here. There are some statistics that we use, this is called the standard error. A small standard of error is more desirable than a large standard error. Let me apply that to a teacher effect. We're going to get a number; that number is a teacher effect or a point estimate. That number -- and then we get the standard error that tells us, well, how certain are we that this is where the teacher's ranking really is? A small standard error tells us, the variability is pretty small. A big standard error means we've got a lot of uncertainty. It's not very precise. So what are we looking for? We're looking for a model that yields with other things being equal smaller standard errors. This is what we want to see. These are data that we're about to show you.

Why do you care? Well, a standard error tells us that the estimated teacher effect is more precise. You don't want to estimate teacher effects with a lot of uncertainty. American Court Reporting 850.421.0058
that the black dot would be over to here to the  
left indicating that on average it has a smaller  
standard error -- by the way, this is math,  
grade seven math. Two of the black dots over  
here to the left indicating that the model has a  
smaller standard error relative to the  
standard error of other models, these are  
standard errors of the teacher effects.  
Then we want to see smaller spread. We  
don't want this box to be big. That's not  
desirable. That means on average of small  
standard error but there's a lot of variability.  
What we want is the black dot to be on the left  
and that box to be smaller. That would be  
desirable. So what do you observe? Any  
reactions to the teacher effects standard  
errors?

MR. LeTELLIER: Model 3C, 3B, and 3A all  
are about the same, the spread and they all have  
similar low error range. That's what I would  
quickly say.

DR. DORAN: That's a good observation. I  
want to go to another person, but I want you --  
I'm giving you an assignment. I want you to go  
to your cheat sheet and I want you to find any  
American Court Reporting  
850.421.0058

Let's do something real quickly here. We  
can just tell you an observation that you've  
already made. Suppose we were to do the  
ranking, which one took the smallest standard  
dots. You can see that this one is smaller than  
this one on average and this one is smaller than  
the others. We're going to do this as we go  
in that column precision, now you can make a  
judgment about the terms of their precision.  

MR. LeTELLIER: So I'm going to suggest  
that we're going to do this as we go  
forward with today because by the end when we  
got to the last criteria, we don't want you to  
be overwhelmed with "I forgot which model was  
good at precision, which model was bad" --  

MR. LeTELLIER: We are looking at reading  
and math for all these categories?

DR. DORAN: Yes, for all these categories.

MR. LeTELLIER: So I'm going to suggest  
just because -- I think it was Pam or Anna had  
American Court Reporting  
850.421.0058
mentioned about if we're bound to use the same model for both and Juan said no; maybe we should put a slash here so that we can look at both.

DR. DORAN: Yeah, put a little line through it.

MR. LeTELLIER: This way we're evaluating for reading and math, and at the end of the day it's the same, great, but it'll be easier to remember, I think.

DR. DORAN: Yeah, I wish I had given you two sheets.

MS. FEILD: Can I ask -- when you ran them going back to the models with the two year lag ones that seemed to have better precision, did that include every 7th grader in the state regardless of whether they had two data points or three data points, or was it only --

DR. COHEN: No, this only included the ones that had both data points.

DR. DORAN: Let's say something about this. You only put a student for two data points --

MS. FEILD: No, no, we have -- these models -- the top models all yielded two lags. That means they would have had three data points.

American Court Reporting 850.421.0058

123  

DR. DORAN: Correct. The correct score for the --

MS. FEILD: In other words, your fourth grade model for that component would never have two lags.

DR. COHEN: Yes, we didn't estimate it for the fourth grade for exactly that reason.

MR. COPA: That's a great question and it leads to the policy implications of you need to -- you know, considering precision but also what are the consequences of selecting a model that requires two lags.

DR. COHEN: Right.

MS. FEILD: Right. That's why I'm asking -- but the first question is what was plotted before.

DR. COHEN: It included only students for whom you had both data points, and there's nothing preventing me from making a recommendation that says use two data points if you have it; if you don't, don't.

MS. FEILD: Okay. So if we were to choose a model that looks at two lags -- 3A, 3B, or 3C -- let's say we all decided on 3B right now, whatever. Does that mean that we would only generate -- for a teacher we would only generate data for the kids who have two lags?

DR. DORAN: No, no. We can make a decision so that when students have two prior test scores, there can be a decision, a policy decision. When students have the two available test scores, you use them if and if a student has perhaps only one of the two, then what we would do is we could put in what's called just a code that would indicate that one of those two scores is missing and only use one of the scores. Of course, if the student doesn't have anything, then you can't use them at all. But I don't think we can use them -- but you can make a decision.

Use two where available and when it's not available use one of the two, or maybe you just use the most recent of the two.

MS. FEILD: So if we were to use that combination meaning however many it's one set. If you have two, three years with current, you use three. Do we know how that affects the precision of the model?

DR. COHEN: Not -- there's a standard error for each individual teacher, each teacher's American Court Reporting 850.421.0058

125  

DR. COHEN: Yes, we have a little less error around their estimate would be a little bit larger.

MR. LeTELLIER: So that wouldn't be as precise.

DR. COHEN: Yes, we have a little less precise estimates for teachers and a lot of the kids in the classroom who didn't have the extra data. You have more precise estimates for teachers who did have the extra data.

MR. LeTELLIER: So what would be the teacher implications on that one.

DR. COHEN: Well, that depends on your classification which we're going to get to at the end of the day. But you have variability in teacher classification that are much more important than that right now. The teacher teaches a lot of students; they're going to have a much more precise estimate than if they teach a few students, and there are big differences in teacher precision. If we page back a couple of these, let's see -- you mentioned Model 3B, it's right here. You have some teachers -- 25% of your teachers have standard errors of 14 American Court Reporting 850.421.0058
went all the way down here, that would be good.

If I had this and I had a big wide bar that
exactly the variability that we're interested
what the graph suggests.

So you want more teachers down at that end of
the scale which is the reason we display it that
way so you can see where you have a lot of
outliers up here.

No one asked --
No, no, there are other
factors. Two lags brings additional precision.
There are other things that yield more precision
as well. So the standard errors of the teacher
effects depend on a number of different things.
So, for example, it depends on the number of
kids within a teacher's class. It depends on
the homogeneity of the students within that
class. It depends on the number of lags that's
used. It could depend on some of the fixed
effects. So there are other factors, yes, those
grade four teachers will never have a prior lag
as the grade five teachers and up would. But
that's not the only thing that determines what
the standard error of the teacher fixed effect
is.

Jon, you wanted to say something about the
fixed effects, right?

Well, we very quickly started
focusing on the covariate model. The simple
differences model, Model 4, displayed on the
American Court Reporting
850.421.0058
top, you see the standard error is there. I'm eyeballing this. The standard error is around 40 or 50 points. If you'll recall from this model, the estimated effects were also larger. The typical effects were estimated to have larger magnitude, either negative or positive, but larger magnitude. So fifty points relative to the size of the effect being estimated is not as bad. I want to make sure that we don't ignore this model unless the committee wants us to ignore the model.

I think the -- you can see why -- let me back up a bunch of slides and show you this model is finding bigger typical effects. It's because we're estimating something different. It's estimating deviations from a different line. Let me go back there and show you that again.

Remember the red line here, this is a simple differences model, and this is the one that shows you what is typical, what is currently typical. The magnitude of the effect, a typical student -- a teacher teaching only typical students with typical growth right here will appear to have a big effect. I don't know.

DR. DORAN: That's correct. So Model 4 is the simple differences model and all of these are the covariate adjustment models, and Models 1 and 1A are the ones that have the teacher effects and anything with a 3 -- here's an easy way to remember it -- anything with a 3 is a 3-level model. It has students, teachers, and schools.

MS. FRAKES: Thank you.

DR. DORAN: Jon, is that why they were called Model 3?

DR. COHEN: I'll exclaim yes but I might not be telling the truth.

DR. DORAN: I actually just thought of that now and I thought, wow. Just because the way we write them and the analysis specifications.

DR. HOVANETZ: I will have to say, though, we did revise them to have simple differences, that's the reason why a two tiered model is 2A and a three tiered model is 3A.

MS. FRAKES: And the others are the.

DR. DORAN: Yes.

MS. FRAKES: And the others are the covariates?

DR. DORAN: That's correct. So Model 4 is the simple differences model and all of these are the covariate adjustment models, and Models 1 and 1A are the ones that have the teacher effects and anything with a 3 -- here's an easy way to remember it -- anything with a 3 is a 3-level model. It has students, teachers, and schools.

Remember the red line here, this is a simple differences model, and this is the one that shows you what is typical, what is currently typical. The magnitude of the effect, a typical student -- a teacher teaching only typical students with typical growth right here will appear to have a big effect. I don't know.

American Court Reporting 850.421.0058

DR. DORAN: Well, it works out to be a beautiful heuristic device. If you have a 3 next to the model, it has teacher, school, and student effects. So that's why I prepared that little summary sheet for you so you could follow along because we're going to refer to these models by number throughout the course of today and tomorrow.

Yes?

MR. LeTELLIER: Real quickly, the regression line that you just --

DR. DORAN: You want to revisit it?

MR. LeTELLIER: I don't know if we need to revisit it. It's on page 11 you just showed, but it shows the slope is straight, and my question is if you were going -- if a kid one year was 1,400 and the next year you're expecting him to be about 1,450 or something like that, if the kids in that area had a specific number that you're expecting them then would that line be straight or if you magnified it would it kind of wobble as it's going up?

DR. DORAN: Well, the simple differences model forces the slope of that line to be fixed.

American Court Reporting 850.421.0058

how big that is on the scale. That's probably 200 points on the scale down here just for teaching typical students, so there's that big gap. Up here you also have that similar sort of advantage, just not as a disadvantage in this particular range. You see more variability. See estimates of effect that have a greater magnitude because of deviations from a different set of lines.

Does that make sense? Okay. So now we can leave this model for as long as you guys want to leave it.

Next slide.

DR. DORAN: So moving on to the next criteria, we don't want to evaluate any model in isolation. Personally, you might care about precision more than you care about personal. Those are your own opinions. Sam's going to facilitate a conversation later on where you get to express which of these criteria you place more weight to. This is up to you, all right, but don't evaluate (inaudible) -- you get a model on one of these criteria. So why don't we move on to the next one unless someone wants me to stop here and explain more or have larger
at one, okay. The covariate adjustment model takes all of those idiosyncrasies across and finds the line, the slope of that line, that best fits the data given some of those wobbles because you do expect some wobbles. Now you can do other things statistically to account for some of the curves and the data and you can fit a whole bunch of different things, but that line finds the best fit of that line through that cloud of data.

MR. LeTELLIER: Okay. I guess I was having a hard time with that because it just seemed like we're making that slope and now we're just -- all kids are following this where you may have one area of growth that may be higher or lower, and so we can't really account for that precision.

DR. COHEN: Well, we do in fact, and the other model where we could -- let's go back to that. We don't have a graph on this, but with the help of a graph we do have a bit of imagination as an interpretive dance. All right. Ignore the red line, ignore the man behind the counter, ignore the red line. We'll go to the green line, it's just a straight line.

American Court Reporting 850.421.0058

Now let's suppose we include this covariate a little flag which we did, a flag that says this is an ELL student, a student who is in one of his first two years of ELL. There will be a separate line for those kids. It will be parallel to this line and in fact is parallel to this line and a little bit above this line. Then we had another dummy variable for students with autism spectrum disorder. I'm trying to remember; I believe that one actually had a positive intercept, also. That will be another parallel line that is a little bit above this line. I don't try to remember all the details.

A kid who's been absent for 60 days will have a separate line that's again parallel but dropped this much. Katie's been absent for, I don't know, more days than that, but as per 100 days might have a line down here. So the covariates actually account for that sort of -- those sorts of differences among students given their measured characteristics. It doesn't try to fit just any individual differences in the data.

MR. LeTELLIER: Right, I understand that. I remember that from last time, if you're looking at the X and Y access the kids that are American Court Reporting 850.421.0058

...
MR. LeTELLIER: Yes, that was much more eloquent.

DR. COHEN: I suppose we could certainly do that.

MR. LeTELLIER: It doesn't make sense because what we were discussing last time, and I'm just throwing this out here, because what we were looking at, and I remember clearly conversations because I remember the number 140 offhand, and we were talking about students next year, we were going from a baseline, those students would be -- the teacher would be assessed according to where the general kids at 140 were the next year. And with a straight line with that big of a spread, there's no way to do that, and in the center of the line obviously as you moved the slope there's less movement and the outer parts move more.

DR. DORAN: Can I ask for a clarification? Are you asking what would happen if we took a particular range of the X axis here of those kids that were in that range what the expected score on the Y axis? Is that what you're asking?

MR. LeTELLIER: Yes.

American Court Reporting
850.421.0058
year we'll see they're at 1,100. We're going to see those two numbers. So when we're looking at outcome and we're saying where a teacher should have their students, we can visibly see that the general student that was at that level should be at such and such level.

MR. DORAN: That's what this is doing, that's what this is doing. This is showing for any given -- for any kid who had the same prior test score, what are they expected to do in the subsequent year? That's exactly what this line does.

DR. COHEN: There is a linear assumption here that that data goes in a line. He's saying couldn't we do something non-parametric where the line and plots go wherever it wants to.

MR. LeTELLIER: Yeah, I understand this plot. I see where the kids were the first year or the second year, but life is not linear.

DR. COHEN: It's often more linear than you think, but let me -- I have a question over here.

MS. BROWN: Well, I just want to see if this is where you're trying to go because it isn't linear when you look at each case in American Court Reporting 850.421.0058

1 point, and that's absolutely true. The purpose of this is to look at a large set so that we see on average because those -- if I'm understanding this correctly, all the points are actual performance, and that line is not a forced line. It's not a -- we have determined by some calculation that the line should be here because that's what we should expect of students; the line is actually where the average falls. So in a sense even though it is linear because when you draw a progression line through a scattered plot of points it will be a linear line, but this is actual performance and it's attempting to show the average over a large set of data.

The problem is going to go back to kind of where you were talking before. What if it's only that one kid, and that one kid really makes a difference for that one teacher?

MR. LeTELLIER: Well, yeah, let me just draw this.

MS. BROWN: Go, John, go.

MR. LeTELLIER: I'm just going to use three numbers here. What I'm saying is this: If we draw a line that goes like this, and I totally get that; what I'm saying is what if the kids American Court Reporting 850.421.0058

1 that were right here on average actually score here; that would make it here. The kids here on average score here and the kids here on average score here. When you connect those points these are still average. Again, that's not what I'm saying.

DR. COHEN: Let me show you what you would see. If this were the case you would see that on that graph.

MS. BROWN: You would see a plot line.

DR. COHEN: Let me draw you what you would see. If we try to defend it around this line, you would see a scatter plot that looks something like this.

MS. BROWN: So the scatter plot would have to be shaped --

DR. DORAN: You see that plot in the residual and you don't see that, you don't see that here.

DR. COHEN: Let me see if I can -- I have some data --

PANEL MEMBERS: (Over-speaking.)

MS. BROWN: Can I just say that in like common terms?

DR. COHEN: Yes.

American Court Reporting 850.421.0058

144

143
would see the scatter plot like that, okay? Now we compute what's called the residual. The residual was the difference between the kids' observed score and their expected score.

We did many, many analyses to look on those residuals to look at whether they fit particular patterns. They normally distribute it, how did those fitted values plot -- the fitted values plot against some of the residuals? Those strange trends would have shown up in our analysis of the residuals, and we would have thought we need to go back and fit cubic trends, quadratic trends, do some local smoothing, you know, use different polynomial terms to account for the bends in the data. We actually did do something like that back at the lab to see if those models fit the data better than these straight line models and they didn't. Not only did they not, but in all of our analysis of the residuals that we do that we're not presenting here, but if you want to see we can show you some of these things, we don't see those strange patterns of the curvature that we would get to. Perfectly reasonable what you're asking to look at whether that's true or not because if we

---

146

1 all models are wrong, some models are useful.
2 We can't fit a model that fits, you know, into every single one of these points. That would be what's called a saturated model and it's not a reduction of the real world. So we have to investigate which model is the best reduction of the real world data that we can apply and get reasonable inferences from. So that's what we're trying to do here. That's why we're looking at other fixed effects, other covariates and control variates. Adding those terms when they fit the data and then when they're excluded.

147

1 do see that the role is not as linear as these models have turned out to be, because it's just a line, a straight line -- okay, perhaps the world isn't as straight as we made it here. We can test those assumptions, we did do some analysis. I can show you some of this stuff if you want to come over here to the computer and look and see.

DR. COHEN: Just for public record, the world is round.

MR. TOMEI: We're never going to be able to develop the perfect model. You develop the best model you possibly can and then to handle the extreme outliers you have to have some policies or business practices in place, you don't evaluate a teacher, you don't have a minimum end of student population and you don't make high stakes decisions about teachers based on a single year's worth of data. Those types of things will mitigate the kind of concerns you're expressing in ways that the best model in the world would never be able to capture 100% of the time.

DR. DORAN: There's a very famous statistician; his name is George Fox and he said

148

1 value-added models include school effects, some of them include only teacher effects. Let me talk about why that's important and what matters, all right?
3 When we think about school effects and teacher effects, a teacher effect is essentially in the world of value-added modeling the things we think that teacher has done to cause the improved or not improved student levels of achievement. However, suppose there are other things happening in the school that are also cause for students to improve in their learning or not improve in their learning that the teacher is not responsible for.

Suppose there are qualities of the programs initiated by the principal or the curriculum coach or some other advocacy group within the school. There are things that are systemically happening in the school that are a cause or at least part of the cause of student change that are outside the control of the teacher. If you do not include school effects, those changes, those things have to go somewhere; and if you ignore them, they get pushed into the teacher effect. What would end up happening is if you
1 ignore school effects and let's say there are
good things happening in the school and let's
suppose those school effects are real and big,
what happens is they get pushed into the teacher
effects and those teacher effects may appear to
have higher added value rankings than they truly
deserve because they're getting some credit that
the school is doing.

I look really good because things that the
school does that I didn't initiate got pushed
into my effect or vice versa. Things that the
school is doing are dragging down a particular
teacher's effect. All right? So that's part of
the reason that we're interested in looking at
this. Are there things at the school level,
initiatives or programs that are also partly
because of changes in student growth that we
want to account for? When you do that then what
you do is you partition the variability and
growth. We get teacher effects and we get some
of the variability and growth that's due to the
school. Those are estimated.

So what do we look at? That's the
question. Should school effects be included?

What statistic do we look at? We're going to
American Court Reporting
850.421.0058

1 look at the variation in student growth between
schools.

Evidence in favor of a desirable model. If
the model suggests that there are systematic
school effects then the policy must decide how
much to attribute. We don't need to answer that
question today; the question will be: Should
school effects be included or not? I don't
think that's the goal; we're not going to answer
that, but you can attribute none, all, part,
some fraction of that. The issue here is should
we include school effects in the model or not?

Why? As I said, determining if and how
much of the school effects should be attributed
to the school, then also if the school is doing
something that the teacher is not responsible
for and you exclude school effects, teachers may
appear to be higher performing, higher value
added or lower value added than they may truly
deserve.

Let's look at something and think about
this as we get ready for lunch. This is the
only graphic I'll show you before we get ready
to -- I thought we'd finish our slides by 1:00
today. I was way off.
Let's pretend for just a moment that we included school effects but suppose this part of the school effect was really small and this was really big, okay? And this was small and this was small and that was small. That would suggest that even after accounting for school effects, teacher effects are still really big. If that were the case, we might say school effects are not necessarily needed. Now I can't tell you how big the effect of the school should be before you make a decision, a policy decision on whether or not you include it. This is data, these are data that suggest that school effects seem to account for a large proportion of the differences in student achievement. Maybe schools matter.

When we come back from lunch, Kathy, I'll let you close this out soon, we'll explore the observations that we're making about the differences between models and whether or not you think school effects should be included or only teacher effects, and we'll talk about any questions you might have.

MS. HEBDA: Thanks, Harold. A lot to chew on during lunch. We're going to stop the recording and break for lunch now. (Whereupon, a lunch recess was had.)

DR. DORAN: Before we went to lunch, we started with the question of whether school effects should be included in the value-added model, in addition to teacher effects, or only teacher effects. Let's go back a step and talk about how do you do school effects, or what is a school effect? Is it a certain set of variables that you put in? Let me try and describe this in a way that is relatively straightforward to understand what the school effect is. But let me start with just the teacher effect and then we can generalize what the teacher effect is to the school effect.

When we estimate the regression model for teacher effects, essentially what we get is we get an expectation for all kids. We talked about this last time. We said what's the expectation? Remember that regression line? That's the expectation. Every kid in your class, Sam, had positive deviations from that typical expectation. So now we've got what we call a residual. It's a little bit more than just taking the average, but let's just work with this.

So now I've got for all the kids in your class, I take the average of those residuals. They're all positive, which means every single kid in your class for the sake of argument meets the typical expectation. Let's just say they beat that expectation by a lot. Your teacher effect would essentially be about the average of those residuals. They're all positive, you look really good relative to the typical teacher in the state.

Now I've got another teacher and so suppose now I have a group of students and I didn't do as well with those kids as would be expected. Suppose each of the students in my class fell below that typical expectation. Now I've got all those negative residuals and we take the average of those, let's assume, very high positive deviations from the typical expectation. What's the expectation? Remember that regression line? That's the expectation. Every kid in your class, Sam, had positive deviations from that typical expectation. So now we've got what we call a residual. It's a little bit more than just taking the average, but let's just work with this.

So now I've got for all the kids in your class, I take the average of those residuals. They're all positive, which means every single kid in your class for the sake of argument meets the typical expectation. Let's just say they beat that expectation by a lot. Your teacher effect would essentially be about the average of those residuals. They're all positive, you look really good relative to the typical teacher in the state.

Essentially, what the teacher effect is, the teacher effect is the deviation from -- the kids in that class -- is the deviation of the students in that class from the typical expectation. So we know what the expectation is, kids in a teacher's class in many respects -- kids in that teacher's class had in effect a deviation of plus ten. It just means maybe they scored ten points higher on average than the typical student, than was expected, or a negative teacher effect is one where the teacher has -- let's just call it negative ten. That means students in that class performed lower than was typically expected. That's the idea of the teacher effect. It's a deviation from the expectation for the kids in that class. You have positive deviations for all your kids; I
have negative deviation for all of my kids.
2 Now the world is not that straightforward.
3 Your class maybe some of them have positive
4 deviations, some of them did not, so there's a
5 little bit more than that. Again, you still
6 take all of those residuals and we do a little
7 math with those and we come up with, well, for
8 the most part your kids did pretty well and for
9 the most part my kids did not do as well. So
10 the teacher effect is the deviation from that
11 expectation.
12 The same concept applies to the school
13 effect. So essentially what we've got is a
14 group of kids who are within a school, from many
15 schools, and essentially what we're looking at
16 is those kids, how did they fare relative to the
17 expectation? All of the kids in the school?
18 Then kind of in this -- now when we actually do
19 the math, mathematically now what we do -- now
20 I've got this effect. See, all of the kids in
21 this school on average perform lower than was
22 typically expected. Now I look at, well, the
23 kids in Lance's class, they're in that school;
24 how did they perform relative to the school
25 average? So now I've got kids in your class,
American Court Reporting
850.421.0058

they performed higher on average relative to
other kids in that school. Me, kids in my
class, did poorly relative to other kids in the
same school. So now I've got two deviations for
a school effect.
6 I've got an expectation, I expect all kids
7 to learn. Then I've got a school effect. All
8 the kids in the school, how did they fare on
9 average relative to that expectation? The
10 school effect is still a deviation just like the
11 teacher effect is, but now I've got a second
12 deviation, now I've got teacher effects. How do
13 teachers now deviate from that average school
14 effect?
15 How do you do this? It's a statistical
16 thing, all right. We essentially see that it's
17 kind of like the way I described it.
18 Essentially what we get, and again, it's a
19 little bit more than just an average, but
20 essentially what we get are the residuals for
21 all kids in a school. That's the school effect.
22 And for the teacher effect, again, it's the
23 residuals for the kids in a class and we see how
24 far they deviate from that school effect.
25 That's what a definition of a school effect is.
American Court Reporting
850.421.0058

1 It's the deviation. Then the teacher effect is
2 the deviation from that school effect. Yes?
3 MS. ACOSTA: I have a question about the
4 average that we get from -- including everybody.
5 That includes students within that teacher's
6 class so those kids are counted within the
7 average and then separately? Do you see what
8 I'm saying? As well as for all the things we're
9 looking at. When we say we're going to get
10 their regression line for everyone, that
11 includes for example students with disabilities.
12 So we make another line with just the students
13 with disabilities; so they're in the initial
14 line as well?
15 DR. DORAN: That's right. So let's just
16 play it out a little bit. We've got
17 expectations for kids with disabilities and some
18 of those kids are in your class, so we've got
19 observed scores, and we say did they beat those
20 predictions? And they did and there are other
21 kids who don't have disabilities who got a
22 separate expectation for them when that variable
23 is included in the model and did they beat that
24 expectation? They did. So no matter who you
25 taught, all the kids in the class beat their
American Court Reporting
850.421.0058

expectations. Now we've got those residuals and
we'll average those together, say, and they did
better than expected.
4 So where we were going with this is to
5 these slides and asking the question should the
6 value-added model include both school and
7 teacher effects? Part of the reason that this
8 is interesting to look at is because when you
9 look at this here we see that teachers appear to
10 be very different from each other. There are
11 some teachers who would appear to have high
12 value-added and some teachers who appear to have
13 very low value-added. There's a big spread
14 between the teachers. Okay?
15 But when we include a school effect what
16 happens is maybe those teachers appear to look
17 good as a function of something else that's
18 happening systematically in the school. So when
19 you include the school effect the teacher
20 doesn't appear to be quite as good -- almost as
21 good but not quite as good because some of the
22 things that were being pushed out into the
23 teacher effect are now accounted for by the
24 school. So it's going to change and that's what
25 we see in these graphics here. We see that when
American Court Reporting
850.421.0058
1 schools are not included teachers appear to be
2 very different, but when schools are included
3 they seem to account for some total proportion
4 of the variance. It's not tiny, it's some
5 portion. And teacher effects are smaller than
6 when you only have teacher effects.
7 John, you had a question?
8 MR. LeTELLIER: Yeah, the counter would be
9 true as well if you had a negative school
10 effect, then that would change the teacher
11 effect as well?
12 DR. DORAN: It could.
13 MR. LeTELLIER: Okay. I'm not saying
14 that's going to happen, but that could happen?
15 DR. DORAN: It could happen.
16 DR. COHEN: Yes, it does happen. Some
17 schools are better than others -- not better.
18 Some schools show a higher student effect, some
19 schools show lower student effects. Can I say
20 something else?
21 DR. DORAN: Yeah, go ahead.
22 DR. COHEN: This is pure policy decision.
23 We're presenting the statistical data and the
24 data tell you go from policy decision to make.
25 This line here shows you that student
American Court Reporting
850.421.0058

163
growth varies somewhat by school, not as much as
2 it does by teacher, but it varies by school.
3 Meaning that if you take it out of the model,
4 any gross that's common within the school will
5 be attributed 100% to the teachers, and that's
6 fine if that's the policy decision you want to
7 make. Model 1A and Model 1 make that -- it's a
8 policy decision. We've decided that the only
9 thing that matters at the school in terms of
10 student growth is the teacher effects, and so
11 any school level effect that may exist is really
12 just the net effect, the aggregate effect of
13 having all good teachers.
14 So if you want this model, the policy
15 decision you want to make would basically be
16 saying that yes, it's true that more effective
17 teachers are clustered in some schools. So
18 what? They're still great teachers. These less
19 effective teachers are clustered in other
20 schools and, so what? They're just less
21 effective teachers. These models -- they're
22 telling you that there is correlation in teacher
23 effects within schools. Some schools have
24 higher student growth, some students have lower
25 average student growth. Your policy choice is
American Court Reporting
850.421.0058

164
whether to attribute it all to teachers, whether
2 to attribute none of it to teachers, or whether
3 to attribute part of it to teachers.
4 MS. GINN: Question. Then really does it
5 show that without putting the school in teacher
6 impact has allowed fluff?
7 DR. COHEN: That might be a technical term
8 there, right?
9 MS. GINN: You know what I'm saying.
10 DR. COHEN: I can't make the decision for
11 you. This is not a statistical decision.
12 MS. GINN: I'm just asking you, based on
13 what you have there because of the school
14 impact, we'll keep the difference, okay.
15 DR. COHEN: All I'm --
16 MS. GINN: We don't want to touch that,
17 fine. Here's my second question.
18 We keep using the school effect positively,
19 but the school effect can also be negative.
20 DR. COHEN: By definition the models are
21 estimated, half the schools are better than
22 average, half the schools are worse than
23 average. So it's positive or negative. It
24 really has to do with what you believe affects
25 -- if you believe that any common influence in a
American Court Reporting
850.421.0058

165
1 school is 100% student teachers and you're ready
2 to stand by that, choose one of these models or
3 use one of these models that has the school
4 effect back into the teacher effect.
5 MS. HALL: How are the schools being rated?
6 Is it the same thing as the teacher or is school
7 grades or how are we determining -- how are you
8 coming up with that graph?
9 DR. COHEN: Oh, the school effects are
10 estimated the same way. I am going to speak
11 loosely, this is not mathematically how it works
12 exactly, but it is this general idea. It gives
13 us a -- if I were to take these estimates that
14 are only teacher effects and find the average
15 effects within school, I would then subtract it
16 off the teacher effect, put it down here and
17 make it the school effect. So it's kind of --
18 if you leave school out of the model, it's kind
19 of the average teacher effect.
20 MR. MURPHY: If you don't remove the school
21 effect, you have teachers that may or may not be
22 earning that additional portion.
23 DR. COHEN: Yes, it's important that you
24 recognize that as a policy and substantive
25 decision. If nothing statistical models can
American Court Reporting
850.421.0058
166

1 tell you, it's not -- sure, Harold or I would be
2 happy to tell you what the right answer is, but
3 not statisticians and not giving you an answer.
4 This really has to do with what you believe
5 matters in an education system and how you think
6 accountability ought to be distributed between
7 leadership and the teachers.
8 DR. DORAN: This decision would be really
9 easy. Jon's exactly right; this is a policy
decision. Remember, we looked at precision
10 before. Everything is policy decision, slash,
11 statistical decision. But here if we did this
12 and we said that schools didn't account for
13 anything at all, your decision would be
14 relatively easy. There's no need to account for
15 the school effects because they don't account
16 for any other variance and scores. But here we
17 don't see that to be the case. We see that
18 schools seem to have differences between them
19 and that that seems to affect the differences in
20 the teacher effects, right? So there is
21 something that matters systematically between
22 schools.
23 Yes?
25 MS. FEILD: Let me make sure I understand.

167

1 The school effects here is really the same as
2 the teacher, but it's kind of aggregated for the
3 school?
4 DR. DORAN: All kids, yes.
5 MS. FEILD: So what happens is, and I think
6 someone may have asked this question, you may
7 have a school who was rated under our
8 accountability system as an 'A' school, right,
9 based on points? And it could have been
10 proficiency, gains, whatever it happens to be.
11 Now you have this model where the teachers
12 performed okay but not great, but based on the
13 nature of the statistical model where half are
14 going to be below the average and the other
15 half, this school who is an 'A' may have every
16 teacher get a negative school effect on their
17 individual evaluation; is that correct?
18 DR. DORAN: No, it depends. We're going to
19 get to the end of the day -- now, if we use the
20 average as a classification rule, you say --
21 classification is something that we're going to
22 talk about at the end of the day today. Suppose
23 we say that any teacher above the average if
24 it's called a high value-added, any teacher with
25 a below-the-average is a low value-added.

168

1 That's a classification rule and now how those
2 effects relate to the other grades, I have no
3 idea and I don't want to speculate on it.
4 Whether it's plausible that they could look good
5 under one model and not look good under another
6 maybe, but there's no basis for me to make that
7 decision. That's pure speculation.
8 But the issue about classification is
9 something we'll talk about later today; and that
10 is do we have to use the average? No, you can
11 use above the average to create more stringency
12 and go up by five, one, you know, you can be one
13 standard error above that, you can be two
14 standard errors, one-and-a-half standard errors.
15 We can come up with different classification
16 rules and we're going to show you the
17 consequences of what happens when you choose the
18 classification rules at the end of the day.
19 MS. FEILD: I think you misinterpret --
20 DR. DORAN: But how we score grades,
21 there's no way to --
22 MS. FEILD: I understand. I think the
23 misinterpretation is when we talk about a school
24 effect, at least in the original conversation
25 before lunch, I was thinking we were looking at

169

1 other variables that had to do with schools
2 separate from value-added, school grades,
3 accountability, but in --
4 DR. DORAN: No.
5 MS. FEILD: But in essence we're talking
6 about the same thing. Okay. I just want to
7 make that clear.
8 DR. DORAN: That's a good clarification,
9 all right. So essentially, another way to think
10 about this it that we've got -- this is all
11 value-added, all the same metric, and this is
12 everything that we're looking at. So
13 essentially we're going to have some expectation
14 and now I've got -- I'm just drawing -- it's too
15 low. I'm a school who on average beat the
16 expectation and now I'm going to have teachers
17 in that school, some below the school average
18 and some above that average by some amount.
19 It's all based on the same data, even the same
20 model. We don't have other variables like
21 school grades and so forth.
22 Now what's interesting here is that we see
23 a couple of things, schools appear to account
24 for some non-trivial proportion of the total,
25 variation in scores. Non-trivial is not
1 determined statistically. We’ve just shown you
2 the graphs and just based on the visual displays
3 it seems to be a large proportion of the total
4 variance of scores. Teachers are less different
5 from each other when including school effects.
6 What does that mean?
7 When we don’t include the school effect, we
8 have this much variability between teachers.
9 There’s a lot of variability. Teachers are very
ten different from each other, but when you include
11 the school effects part of what made that
teacher look really, really good before gets
13 soaked up by the school effect; or part of what
14 made that teacher look really, really really
15 low value-added before gets soaked up by the
16 school effect. The teachers appear to be less
different from each other under that model.
18 Yes?
19 MS. FRAKES: So when the school effect is
20 smaller, the teacher effect could be greater.
21 So when you’re talking about --
22 DR. DORAN: It would be greater by
23 definition, so as this gets smaller this will
24 get bigger.
25 MS. FRAKES: So when you’re talking about
American Court Reporting
850.421.0058

1 moving your best teachers into your lowest
2 performing schools, I mean, when you’re taking
3 into account as a school this is the school
4 effect that the students are not performing, and
5 then that teacher goes there and she does what
6 she does so well; that’s really going to show as
7 opposed to that school effect, correct?
8 DR. DORAN: That’s correct. The teacher
9 would appear to be very high performing within
10 that school.
11 MS. FRAKES: Within that school because the
12 school effect is lower. I mean, it can work
13 both ways.
14 DR. DORAN: Correct, if you’ve got a bad
15 teacher going into a good school that teacher
16 would have a lower value-added effect within
17 that school than they would in some other
18 situation potentially.
19 MR. LeTELLIER: So could this have an
20 unintended consequence of a performing arts
21 school, which typically we have data that shows
22 that those are higher performing schools, or
23 some sort of a magnet school where, you know, an
24 IB school or something where the kids are very
25 well performing. Will that make it harder for a
American Court Reporting
850.421.0058

1 good teacher to have a good value-added model
2 score because those kids are already performing
3 so high, the school is performing so high, the
4 teacher goes in and the kids are doing well.
5 DR. DORAN: In some respects, yes. In some
6 respects that is almost what you would expect.
7 If those kids are doing well because of things
8 the school is doing then you want to
9 differentiate that from the teacher effects,
10 right? But if you ignore the school effects in
11 that particular case then everything that’s
12 happening systematically in that school gets
13 pushed into the teacher effect, and those
14 teachers may appear to be high value-added, not
15 necessarily because of what they’re doing but
16 because of other initiatives in that school.
17 But now you account for the other things
18 happening in that school, and the teacher
19 effects could -- in fact, they will be smaller
20 as we see here. The teacher now has to
21 demonstrate that they’re doing things above and
22 beyond what’s normally happening in that school
23 to have a higher value-added effect relative to
24 other teachers in that same school.
25 MS. BROWN: I think what is discomfort and
American Court Reporting
850.421.0058
as they're doing those things, they have equal opportunity.

DR. COHEN: Can I break in here?

DR. DORAN: Yes.

DR. COHEN: It's a little more complicated than an extra decision that you have to make. I'll put a little example up here. Let's say we're going to talk about one school. That school has four teachers, and the math is not exactly -- this math does not come out exactly the same, but it comes out really close in this illustration.

So the school effect, estimated according to let's say Model 3A -- school effect is 10, teacher effect is 20 minus 20, 10 and minus 10; so they're equal. They have an average of zero around the school mean. If we were to ignore the school effect and just estimate it for the teachers and say everything is attributable to the teacher, those scores under Model 1A would be 30 minus 10, 20, and 0, right. So this is attributing all effects to the teacher, so we just assume that any school effects are just the result of an aggregate teacher effect.

If we take 20 minus 20, 10 and minus 10, American Court Reporting 850.421.0058

that's attributing 100% of the school effect, of whatever is common in the school to the school. So great principal or a bad principal, it's all on the principals. But what we did down here is we say, well, we decide we're going to split this half and half. We're going to say half of it is due to the fact of the school and half of it is due to the teachers themselves. Well, then we wind up 20 minus 15, 15 and minus 5.

The thing is if we estimate it with the school effect, this decision is explicit. Do I add these things together and get that? Do I partition it partially and do that or do I attribute it 100% to the school?

MS. BROWN: However, if you look at that differently and you use your same numbers, I think the question around the table is if school A has a school effect of 10 and school B has a school effect of 5 yet the same teacher effect of 20 in each of those schools -- or maybe I'm backwards, but the point is that is if we went that way, is it possible that in the long run teachers at the school effect 10 have the potential to have the higher scores? And if there was a lower school effect, would those
178  
1 this is part of what you're getting at. Suppose 
2 I had two schools or two teachers and suppose 
3 some teacher had a -- suppose at the school A 
4 it's typical for the students to score 20 points 
5 above the state average. Let's just use that 
6 term, it's typical for kids in school A to score 
7 20 points above the average. Now suppose that 
8 that teacher, a teacher within that school has 
9 kids who only score 20 points above the average. 
10 That teacher didn't do anything above and beyond 
11 what's typical for students in that school, 
12 right? 
13  
14 Now let's take another situation, another 
15 teacher, who's in a different school. In that 
16 school, let's just say it's typical for teachers 
17 to score 10 points below the average. That's 
18 typical for that school. But that teacher, a 
19 teacher in that school scores 20 points above 
20 the average, the same as the other teacher, 
21 okay, in the different school. That teacher 
22 would appear to have a higher value added 
23 because that teacher is doing something that's 
24 very different than what's typical in that 
25 school. Do you see? So while two teachers did 
26 the same thing, this teacher over here is doing 
27 American Court Reporting 
850.421.0058  
179  
1 what's typically expected or typically observed 
2 for kids in that school. There's nothing 
3 exceptional happening there, whereas this other 
4 teacher went away above and beyond what's 
5 typically observed for kids in that class. 
6  
7 Now while two teachers have the same -- 
8 both have kids that performed at the same level 
9 given that they're in different schools this 
10 teacher did what's normally observed -- or did 
11 what's normally observed in the school, so it's 
12 not regarded as being high value-added because 
13 that's what's typical for that school. In that 
14 instance you could have teachers who do the same 
15 thing with their kids, but this teacher over 
16 here is doing something that -- that's what you 
17 observe with any other kid in school. So what 
18 makes that teacher particularly high value-added 
19 when the school effect is included? When the 
20 school effect is not included, both of those 
21 teachers would appear to be exactly the same. 
22 They both beat the average by 20, right? But 
23 unless you know something more about the context 
24 of the school that they're in, well, this goes 
25 to -- that's what you typically observe because 
26 there's other things that the school does to get 
27 American Court Reporting 
850.421.0058
MS. KEARSCHNER: That's my question really.

That's what I think I was trying to get at. Is there -- could there be any correlation with the school effect in those kinds of populations?

DR. DORAN: Let's take an example and answer the question. Could there be? Suppose we have a very high performing school and suppose we have a school where there are a lot of students who come in with very high scores in the intake, right? Let's just say all of the kids are clustered in this particular area.

Now, can that school still show a positive effect? Well, one of the things that we saw was there is pretty significant scatter around that expectation line even at the high end of the score distribution, right? It resulted that there was virtually no scatter around that point, we would observe that students for whatever reason weren't deviating much from the conditional expectations of the high or low end of the distribution. It's plausible that being at the end of the extremes, very low end of the distribution, could limit your ability to be classified as low value-added, or being at the very high end of the distribution if you have a cluster of kids who could be there, it's plausible. The degree to which it happens, I just don't know because we don't subset the schools when that happens and see if there is a particular consequence. But to the degree that would play out in the real world, I'm not sure is a huge concern.

DR. COHEN: This really is a big and important issue in the accountability system.

You need to think about it. Unfortunately, it's not a technical statistical issue. If it was, you could ask Harold what the right answer was and he'd tell you. But statistically when you estimate the school effect, and again I'm speaking loosely but it's pretty much true; when you estimate the school effect the teacher effects are going to have the average of zero in each school. So that means that all schools would have some positive teacher effects, some negative teacher effects. The school effect will be either positive or negative. It's around that school.

Unless you believe that teachers are distributed evenly across schools, every school:

American Court Reporting
850.421.0058
has an equal mix of better and worse teachers, high value-added and low value-added teachers, you probably wouldn't want to attribute everything common in the school to something other than the average teacher effect, right? So on the one hand you want some of that to count for the teachers because you don't believe that teachers are equally distributed across all schools. On the other hand, to attribute it all to the teacher, you may want to do that. To attribute 100% to the teacher, then you would say things like school leadership doesn't matter. If you've got great teachers in your school or you say a principal has this effect on student learning through his effect on the teachers; he or she makes the teachers effective, and that's effective school leadership. So it's really a policy decision. You can estimate the model with the school effects and then add the school effects and teacher effects together like I did on the board here and wind up with something very much like this. In that case, you'd be making the decision that I'm going to attribute everything to the teacher-effects, and that's fine. If that's how you believe it works and that's how you want to attribute it, that's fine. That's how you do it. If you agree now that you want to do that, you're probably better off with one of these models because that's done automatically and you don't even have to mess with it. But if you think that whatever is common in the school should be shifted and shared a little bit between school leadership or whoever and teachers at the school. You estimate one of these models, but you think real hard about whether you just take this teacher effect or whether you add in all or part of the school effect for each teacher as well. MS. FEILD: Is the school effect affected by the differentiation that we had earlier on the growth where a child who's sitting in two reading courses has a higher expectation for growth than one who is not? And the question would be on a low performing school where let's say 90% of the kids are being double-dipped versus not, is that going to negatively affect the school effect because there's a higher growth expectation?

187

1 maybe a different effect. Gina?

2 MS. TOVINE: You've kind of already answered that, but one of the things I was going to ask, since you've run the data for every one of those models and you have all that information, is it possible for us to see at some point like the simulation, like see teacher A and teacher B and how it would actually pan out for that individual teacher for each one of these. But you just kind of indicated how that would actually be translated over to the value-added model.

DR. DORAN: We had a slide that I think we took out and perhaps we can recreate the slide and put it back in. I think the question you're asking is, are there -- when you have the school effects and the teacher effects, do thinks change for the teachers' classification relatively large? The answer is no. One of the plots that I had previously was it showed the relationship between teachers under every single one of the models, and what you see is virtually a perfect correlation between teacher effects and models 1 and any of the models 3. What that suggests is that however teachers are classified...
under this particular model, the teacher only
model, they remain similarly classified under
this one.
If they have high value-added effects under
one model, they would also have high value-added
effects under the other model. We could
recreate that and show that to you so you could
see there's virtually a very strong linear
relationship between the two.
One of the things that might have been a
cconcern is what if the correlation was zero or
even negative? Would it contain the teacher
effect under this model and look at its
relationship to the teacher effects under this
model, and suppose the correlation became
negative or zero, meaning they flip-flopped or
there's no relationship? That doesn't happen.
We know and we can show you that the correlation
between the teacher effects under all of these
models is very close to one.

Yes?

MS. BROWN: I just want to throw something
out for thought, politically correct or not.
But sometimes -- you know, we started our last
discussion when we were here face to face
American Court Reporting
850.421.0058

talking about wishing we could incorporate
things that we don't measure, like parent
involvement, homework completion, things like
that that we have no data for. When we look at
something like a school effect, even though the
residuals are calculated based on what's in the
model, we're still getting an aggregate of that
school. So in some ways we're getting the
hidden variable of that which we wanted to
measure but we can't and we might want to think
about that.

DR. DORAN: We're going to hire you because
that's exactly the right way to phrase it.

MS. BROWN: Thank you.

DR. DORAN: There are things that are
unknown in --

MS. BROWN: I have a very nice fee.

DR. DORAN: That's exactly why. There are
things in that school, parents who do extra
thinks for their kids or who don't do as much
for their kids, or certain policies or practices
or neighborhood effects, or peer effects or
something --

DR. COHEN: There may be, but then this is
really -- there's a reason that there's a
American Court Reporting
850.421.0058

stakeholder committee involved in this process
because what Harold is expressing now is
opinion. He believes this to be -- he believes
this to be true, the -- but the stakeholder
committee here because it's your beliefs about
whether those things are affecting the teacher
effect. It's your beliefs about those things
that should be driving the day and driving the
decision.

On the other side, if you take the school
effect and attribute it entirely to the school
then the average teacher effect within each
school is going to be approximately average,
right? So there is a whole continuum of choices
you could make. That choice depends on what you
can believe; you're the stakeholder. You guys
represent the folks who are going to be affected
by this, whose children are going to be affected
by this, whose staff are going to be affected by
this. You need to figure out what you believe
affects student learning in the school.

MS. FEILD: But speaking from the political
perspective, if you're selling this model to
teachers and you're saying to teachers that it's
not just the math or reading teacher that
American Court Reporting
850.421.0058

counts, but everything else that everybody else
is doing is contributing to the achievement of
the student, then I think you have a better
chance of having a school-wide attempt to work
as a team opposed to -- I mean, it's a
perception.

DR. COHEN: A lot of companies decide that
they'll compensate their executives partly on
overall company performance and partly on the
performance of their unit; so if you estimate
school effects and your compensation system is
compensating teachers partly on overall school
performance and partly on their -- ability --
that's allowed under the current legislation
then yes, you've estimated the school effects
and you've made some decisions on the back end
and then --

MS. FEILD: Wouldn't you say that when we
continue to add additional assets to this model,
when we add a APIBAICE (ph), all of that, then
by having that school effect it allows you to
combine all these pieces, and so the effect of
all of these different assessments could
contribute to a student's or teacher's
performance. Is that going to happen you think?
American Court Reporting
850.421.0058
No.

DR. COHEN: I don't know enough about all the things that are going on.

MS. KEARSCHNER: It's actually required by law that there's pieces from parents, from students. There's all these other things that have to go into the evaluation. That might be the other 50%, but it could certainly be incorporated as part of this school effect. I mean, that's another --

MS. FEILD: I was thinking when we add staying to the achievement side we eventually should be looking at models for, let's say, advance placement scores or international baccalaureate, so if those get tied into an overall school effect then again you're sort of tying in all of the achievement results.

MS. BROWN: Technically, if you tie all those different assessments together at a school level then it's all going to play into the school effects.

MS. FEILD: The teacher may be teaching English in 10th grade, AP English in 10th grade, and that student is taking the FCAT 10th grade reading test and the AP English test, right?

American Court Reporting 850.421.0058

And in theory at some point I would assume at some point that teacher may be measured on both those assessments, both the FCAT and the advanced placement, and eventually there will be more than that.

DR. DORAN: I think this is an interesting conversation. Now, remember, the teacher effects here are essentially the unobserved thing that is happening after we control for all the stuff that we think is true, you know, the difference between kids, there's some deviation from the expectation, and there's a question mark about why that happens. In the model we assume that's because of the teacher and the school effects is kind of the same thing. There's un-observed random variations that happen at the school level that may be due to other things, and that gets called a school effect, and then teachers have to deviate from that in order to have high or low value-added.

Yes?

MR. TOMEI: You talked about the consistency of the models when looked at from individual teacher outcomes, and that's using the data you created so you know that you're dealing, I assume, or is that dealing with --

DR. DORAN: Real world data.

MR. TOMEI: Okay. Here's my question.

There's been a fair amount of research done historically looking at different types of student growth models, including value-added models. The work that was done previously in Florida among the things that were done was they looked at how those models categorized teachers into quartiles, which is interesting because that's essentially what we're going to do under the legislation now; and then they looked at the stability of teacher ratings over a number of years and found way more variance than what you would logically expect to occur. At any point in time, are we going to know how these models evaluate real teachers from year to year over time and the consistency there because I think that's another important attribute that we need to consider.

DR. DORAN: I think Jon and I were the first people to write a paper on this about eight or seven years ago on the volatility of teacher effects that we observe over the time, and the paper we called it From Saint to Sinner.

American Court Reporting 850.421.0058

Pattern, we saw teachers that were sort of bouncing around. That's a well known effect observed in value-added models. When you estimate teacher effects on a limited subset of data, you're going to get an estimate and if you estimate it again the following year you're going to get a different estimate. There are some teachers that will switch from low quartiles to high quartiles, and part of our classification that we're going to talk about today in terms of how you take multiple teacher effects and possibly use them, it may be -- we'll show you -- when you use one teacher effect the probability of being missed by (inaudible) is very big because we don't know. There are swings.

By the term, mis-classified, what I mean is if you're classified as a high value-added, the probability that you're really low value-added could be very high. But as we add additional years of data, the probability of that misclassification goes down. We're going to actually show you data that shows exactly that in terms of what we would anticipate, and that is something that you should look at as you American Court Reporting 850.421.0058
collect more data. We would expect more stable estimates as you use more data.

* * * * *

(Whereupon, this concludes Day 1, Volume 1. Volume 2 continues without interruption.)

American Court Reporting
850.421.0058
effort

E

ear

early

second

down

doubt

drag

dropped

due

dumping

early

earby

earnings

American Court Reporting
850.421.0058
56 of 67 sheets
P-12 [1:1] 46:10
pocket [1:2] 5:21, 48:8
page [1:3] 30:12, 121:15, 122:25, 133:15
paid [1:4] 51:8
pain [1:5] 29:3
Pam [1:6] 4:18, 121:25
pan [1:7] 189:8
panicking [1:9] 82:8
paper [1:10] 190:22, 196:25
par [1:11] 51:19
parametric [1:13] 142:15
paramount [1:14] 78:18
parent [1:16] 191:2
parsimony [1:18] 34:6, 87, 84:14
Parsimony [1:19] 87:24
paring [1:20] 15:4
partially [1:24] 175:13
participating [1:25] 3:3
participation [1:26] 224
Particularly [1:29] 68:13
particulars [1:30] 78:25
partition [1:31] 150:19, 175:13
partner [1:34] 98:6
partners [1:35] 10:17, 56:16
partnership [1:36] 139:18
pass [1:37] 36:24, 72:1
path [1:39] 14:19
Pattern [1:40] 197:1
pay [1:43] 78:16
paying [1:44] 79:15
PD [1:45] 79:25
peer [1:46] 191:22
pen [1:47] 97:19
People [1:48] 69:2
The text is not clearly visible or legible. It appears to be a page from a document, possibly a legal or academic text, but the content is not discernible due to the quality of the image.