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 2

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
Variance can be — I'm using the word "variance" that our schools vary from one another, and that another; that's the conclusion? Variance indicates that schools vary from one their ability to impact student growth. Phrase it would be schools differ in terms of from school to school? Growth? Effects, the kids — Effect? Schools differ from each other. Variability in the school effects means that just said is right. The fact that we see for a minute. So if all schools have the same average teacher effect, that variance would be zero; is that right? Dr. Doran: That's exactly right, that there are -- their schools seem to differ and it has a consequence in the teacher effects and whether or not you include school effects is your consideration.

Mr. Foerster: I want to talk through that for a minute. So if all schools have the same average teacher effect, that variance would be zero? Dr. Doran: If all schools have the same average teacher effect, that variance would be zero? Mr. Foerster: I mean, essentially by showing that we have variance in the school effect, are we not just saying that some schools American Court Reporting 850.421.0058 are better than other schools.

Dr. Doran: What we're saying here -- let me answer your question this way. That what you just said is right. The fact that we see variability in the school effects means that schools differ from each other. Mr. Foerster: In terms of average teacher effect? Dr. Doran: In terms of average school effects, the kids -- Mr. Foerster: Okay, average student growth? Dr. Doran: Yes, average student growth. Mr. Foerster: And there are differences from school to school? Dr. Doran: In terms of -- one way to phrase it would be schools differ in terms of their ability to impact student growth.

Mr. Foerster: Okay, so that there is variance indicates that schools vary from one another; that's the conclusion? Dr. Doran: That's right. Mr. Foerster: Okay. If we acknowledge that our schools vary from one another, and that variance can be -- I'm using the word "variance" American Court Reporting 850.421.0058
that student's growth is related to the school student's learning. Therefore, a portion of things within the school that attribute to that use the school effect we're saying there are teacher effect calculation because if we chose that the school effect becomes a portion of that effect rests solely on the school effect, it's there. So it's not that the final teacher's effect if we decided to include it come in that teacher and whatever proportion of school combination of the student residuals attached to teacher and the student component. Let's say we have -- this bar represents the common component of student learning and this affects the unique teacher component of student learning. If we estimate them both together, we can say how much of that common component is due to teachers and so we can add it back in. If we would just take the unique teacher component of student learning then, Sam, you're exactly right; the average.

American Court Reporting
850.421.0058

teachers may still have high value added. This is a complicated process where we still have to navigate.

MR. FOERSTER: But in so doing with the classification rule, to borrow the term I think you've used, you've un-spooled the school effect, unwound it. I mean, you've gone right back to, okay, well, then that teacher effect is actually in terms of student growth accomplished by this teacher would be this number and that's what we want to look at.

MS. BROWN: But let me clarify because now I'm getting a little confused and I want to make sure I'm right. The final teacher effect is a combination of the student residuals attached to that teacher and whatever proportion of school effect if we decided to include it come in there. So it's not that the final teacher's effect rests solely on the school effect, it's that the school effect becomes a portion of that teacher effect calculation because if we chose to use the school effect we're saying there are things within the school that attribute to that student's learning. Therefore, a portion of that student's growth is related to the school.

American Court Reporting
850.421.0058

would be average in every school whether the teachers in that school or average, above average, or below; the average would be average.

So this common component should probably at least partially or maybe fully attributed to the teachers in the school to move them. So it's kind of a sliding scale. You can take some or all of the common component, plus all the teacher component and use that to calculate your teacher effect. Then when we think about it that way, I think it becomes --

MS. BROWN: My teacher effect based on students' growth that are attributed to me and a little bit partially based on the overall scale that also helped contribute to my --

DR. COHEN: Yeah, yeah, as a teacher me and my colleagues are contributing to this common component.

MR. LeTELLIER: You know, we've spent a long time just on this and from what I understand and just listening and what I'm thinking myself, that's hard to grasp. Here's the scenario and I think this would wrap it up.

If I'm working just as hard at one school and just as hard at another school, could the
school effect actually make it so that I would not have as high a value-added model at one school versus another even though I was working just as hard at each school?

DR. COHEN: That is exactly the question that you want answered, and the answer to that is that it depends on what you believe moves student achievement. That's not something we can give you a statistical answer for. It depends -- this is really -- if I knew what caused student achievement, I'd write a book and retire and all that.

MR. LeTELLIER: Okay, but with what you have with those models, as you increase the school effect you decreased to use a word you used before in another graph the spread of the potential of what a teacher could be effective as, correct?

DR. COHEN: Well, you go back to the old language. No, as I recognize the common component within school of student learning, I acknowledge that there is less of a unique teacher component to it. However that common component is due to my actions as a teacher is the decision that -- it's going to depend on American Court Reporting 850.421.0058

what we believe to be true.

MR. LeTELLIER: So does that consequently from school A to school B, same teacher goes from the same school, if this was working out totally equally, that same teacher that was working hard in school A goes to school B; they should get the same value-added model effect, but --

MS. BROWN: Only if they have the same population of students and the same demographic and --

MS. EDGECOMB: That's the key.

PANEL MEMBER: Right.

MS. BROWN: Because working hard is relative to your belief system of level of effort and --

MR. LeTELLIER: Yeah, I'm saying doing what you need to be doing as a teacher and what we're basing this on is we're saying -- take the kids that are all scoring 96's, we'll just say 96 out of 100. Once you get up to that point, it's very hard to move a kid. So that's obviously a student teacher level.

DR. COHEN: I understand what you're saying and I understand your frustration. So let me American Court Reporting 850.421.0058

give you two examples of world views, two different belief systems. Suppose I believe that school leadership is of primary importance. So anything that any of that component is due entirely to the principal, all right. Then under that situation -- and let's say you go from a school with a great principal to a school with a lousy principal, right? Under that scenario, if the whole common component is due to the principal then you want to completely separate the unique teacher contribution from the common component, and that's a situation under which you doing the same thing with the same group of kids is going to get you the same value-added score. That's one world view that the school leadership is causing that common component.

All right. Now let's go to a completely different world view, and my apologies to any principals in the room -- suppose the principal doesn't matter at all. Suppose that the only thing that affects student learning is teachers, right, and maybe some principals are better at selecting teachers. Maybe some schools are closer to better training institutions; for American Court Reporting 850.421.0058

whatever reason you have some better teachers concentrated in some schools. In that world, if I go from a school with -- let me try to get this right. I've got a school with lousy colleagues, right, and I'm there so the common component is going to -- all right. For the common component, it would be a low score but I'm a great teacher and I come out about average; and then I go to a school -- I'm sorry, I confused myself.

MS. BROWN: What if you take a totally different view and what if you say that you believe that the common pieces are a combination of things, like increased levels of parent involvement, highly involved PTA. Lawrence's point last time, level of resources available in the school, materials, etc., those are things we can't measure. But let's just say that's part of -- if someone believes that that's part of that common component, so then what we're saying is that same teacher, similar effort, but if we say that common component makes a difference then that common component needs to be considered.

PANEL MEMBERS: (Over-speaking.)

Page 208 to 211 of 305
4 of 38 sheets
group of kids that what I did in that school is but now I'm moving to a school wherein I have a let's say the school effect is little to none, already there. If I move to a school wherein up there; it's easy for me because the kids are gifted kids. I'm telling you my scores are way teacher. I'm in a school where I'm teaching you guys that are principals, let's take a for me and it may be that the gentleman and all that? Do teachers have to -- not just speaking for myself -- do teachers need to at this school, school M, a high level of gifted children; so I may be effective but it will be implicated by my population. Over here I'm already a real good teacher, but over here I've got a bunch of sweat hogs if you will that I'm going to need to do something -- PANEL MEMBER: You don't need to -- PANEL MEMBER: Wait a minute. PANEL MEMBER: Wait a minute. you use this model because it depends on how the world really works. Arlene? DR. HOVANETZ: So hand the microphone back to Jon. This is going to be a staged thing. My world view is school effects -- we don't believe that school effect is impacted. DR. COHEN: So in your real world, school effects only reflect the average of the teachers at the school, the average teacher -- okay? MS. GINN: Would you please stand so we can hear you? Thank you so much. DR. HOVANETZ: Another way to think about that could be -- don't yell at me if I get it wrong -- all the student learning that occurs in that school is the result only of the efforts of all of the teachers in the school. That's one. So what we want to know is do you believe that or -- we believe that --
All right. There's only one fair thing to do and that is attribute all the common effects to each teacher.

DR. HOVANETZ: It's Jon's question and I'm trying to rectify this, too, but we keep talking around this issue and it won't give us the actual implication of if my world view is that all of the teachers -- everything that happens in the school is an aggregate effect of what the teachers are doing, and Jon is one school that's got high effects, one school that's got low effects. What is the implication for that teacher's effect? That's my world view.

DR. DORAN: We're getting lost in a couple of things. Let me try and bring us back to something. We're delving into hypotheticals of what would happen if this happened and this happened, and this is going to be a conversation that's going to be circular, and it's going to be very difficult to move beyond this.

Let me try and answer the question. I actually did answer this a little bit earlier. Let me try and state this a little bit differently to try and move this forward.

If you're in school A, in order to be --

American Court Reporting
850.421.0058

and school effects are included or school effects are not included -- what you do in school A in order to have a high value-added effect will be different than what it requires to have a high value-added effect in school B with or without school effects. Conditions change, teaching conditions change, student populations change. When we use terms like if I do the same thing here that I did here, it's kind of a level of abstraction that's really hard for us to attach real meaning to and give you an answer to. So while I like the question and I want to be able to give you an answer, it's only -- we've spent the last hour on this question and we're going to continue to spend the next hour on this question because it is circular.

We can explore various consequences of the if's and and's, but let me bring us back to where we need to be in terms of the policy. Do schools matter?

DR. COHEN: Harold, I think you over-stepped it. I think -- actually, let me try to hijack your example, okay? You two are teachers; please stand up, Mary Ann. You are a teacher who believes that there are a lot of forces out there including principals and families and everything else that cause school effects.

Christy, you believe -- you live in a world where the only thing that affects student learning is you and your colleagues. Okay.

And John, you want to know for Christy and Mary Ann which model would cause them to have the same individual rating, the same individual ranking, whether -- regardless of what school they're at; is that right?

DR. HOVANETZ: Under my world view, what do I look at? A high performing school or a low performing school? In Mary Ann's world view, what does she look like in a high performing versus a low performing school?

MS. BROWN: What would be the range of teacher effects within each world view?

DR. COHEN: Okay. Christy, you're the only thing that matters. If you wind up in a school surrounded by -- well, there's another dimension here. The dimension is model, right? So under which model, right? So let's say we attribute the common component to the school, right? We contribute 100% of the common component to the school leadership; we're not attributing any to the teacher. You find yourself -- so this is this model attributing everything to the school.

The only thing that matters is this teacher. You find yourself in a high achieving school; we're going to under-rate you. We assume you're both great teachers. We're going to under-rate you. You're going to get a lower rating than if you were in a low achieving school, right?

Now Mary Ann, you get exactly the opposite answer. So Christy would prefer to be here where all effects are attributed all and only to the teacher, the common component is entirely attributed to the teacher; that's where she gets the same rating at either one of those schools.

Mary Ann differs in only one respect and that is what she believes about the world is in exactly the opposite situation. This will give her a biased effect because as she finds herself in a school with a rotten principal that's driving learning down, her score is going to be driven down whereas over here it gets subtracted off. So it really is a choice between world views, but they're dichotomous. It's a
continuum. You can choose anywhere in between the two of them.

MS. NOYA: At this point, I'm going around in circles. Maybe I'm incorrect but this is what I want to say. Having done this for so many years of my lifetime, I know principals' evaluations and administrators are also going to be revamped by districts or whatever; teachers' evaluations are being revamped as well.

I don't believe that anything is just without school effects. It does impact it from the top down, bottom up; I don't care how you put it. I've been in low performing schools, I've been in high performing schools. Who you are still will be there, of course. Leadership makes a difference, I truly believe, to support the parents and everything else. But I guess at this point everybody's going to have to pitch in because everything is being revamped. Even administrators' evaluations are being revamped and is going to affect them as well.

So, you know, it's just the luck of the draw. We've been doing this for 38 years. Trying to make it perfect, it's not going to be perfect and there's always going to be flaws.

American Court Reporting 850.421.0058

But I think we've been going around in circles for an hour and have not been moving forward because the impacts we won't know until we start all this, too.

DR. HOVANETZ: But you will know. We can model that. I mean, we can tell you what I the teacher will look like --

MS. NOYA: We need to see that.

PANEL MEMBERS: (Over-speaking.)

MS. NOYA: You might have a poor principal, but then you have a great administrator who drives the school. Principals -- some principals don't run their schools, some assistant principals who are top performing assistant principals run the schools. So it is a lot of variables involved.

MR. FOERSTER: Is it fair to say that we should roll on? I mean, all of us have taken really big swings at this and I think we at least have consensus about what we're confused about. We have a lot of other stuff to go through. Is it okay with everybody if we just keep moving? We'll come back to this; we have to.

MS. GINN: She had her hand up for such a thing. I think apart from the confusion, I think we need a level of comfort that it's going to be fair, what we're doing, the school effects. Jon said if you go to a low performing school and your teacher is graded higher, your teacher effect. We're looking for something that will equalize it; I mean, I'm not using the right words, but for me in my mind I need a level of comfort that we're being fair to teachers in both kinds of schools, and --

DR. COHEN: Nothing's going to be perfect, but we don't want to drive away all the teachers from high performing schools, either.

MS. KRISHNAIYER: And what can help us make that a little more level playing field.

DR. HOVANETZ: Oh, I can't do that but Harold certainly can.

MR. FOERSTER: I'm not so sure we need data as much as hypothetical examples. I mean, just concrete, simple, here's what this would look like, and the thing that I've noticed is missing.

American Court Reporting 850.421.0058
think that our points of view were in contradiction actually; I think assuming that you can break out the school effect completely such that you're net sum at any school is the school average and you've got half of your teachers with positive effects and half with negative, I see that as enormously problematic. On the other hand, I think ignoring that there is a school effect is equally problematic. So where we're going to end up is deciding how we apportion the school effect, and before we can make a reasonable decision about how to do that I think some hypotheticals would be helpful.

**PANEL MEMBER:** Yes.

**MS. BOURN:** What does the same amount of student growth look like as it's impacted by a school effect in a high performing school and a low performing school? And how does that affect my score?

**MR. FOERSTER:** Yes.

**MS. NOYA:** Exactly.

**MR. FOERSTER:** So can we leave it that we'll get some hypothetical examples and pick that up tomorrow at some point when it's appropriate, and then we can roll, we can move on to slide number whatever.

**DR. COHEN:** Let me just point out, you want to know what the right apportion is and --

**MR. FOERSTER:** No, we would like to see examples. I think we all agree that it's going to have to be apportioned. What does that mean?

**DR. COHEN:** Okay, that in and of itself is a huge amount of progress because if you estimate a model like Model 1, you don't know what the school effect is in order to apportion it. You have to estimate this model and then go to the apportioning exercise. So if there is consensus on that you could at least say, okay, we're over here; we have the apportioning.

**MS. FEILD:** Well, then you're saying that you've already decided that your world view is --

**DR. COHEN:** The world view is that it is part of this.

**PANEL MEMBERS:** (Over-speaking.)

**MS. BROWN:** Okay, hold up because I think what really he's saying is what you said last was it's not dichotomous. It's a continuum.

**MS. NOYA:** Right.

applicable, and then we can roll, we can move on to slide number whatever.

**DR. COHEN:** Now, what is parsimony? What are we looking for here? What do we want to know about models?

**MR. FOERSTER:** Does the model control variables without being overly complicated? We could take 30 covariates and dump all those covariates into the regression model, but do you need to? Does that buy you anything statistically in terms of doing a better job in estimating teacher effects?

**MS. FEILD:** That's kind of the question that we're looking at.

**DR. COHEN:** Could you only include five covariates and do a job that is equally as good at predicting teacher effects than using all 30 of those covariates.

**MS. FEILD:** So essentially what we're looking at here is, is the model only as complex as it needs to be? Simple, elegant, accounting for things that are important but not overly complicated to the extent that it becomes difficult to explain, less transparent, and so forth, right? That's the question.

**MS. NOYA:** Is there a statistic we can look at that helps?
helps us understand? Yes, there is a statistic. We're going to look at the percent of current year test score variance accounted for by control variables in the models. Statistically, we call this an R-Square or a proportion of variances. We look at Model 1 from the fixed effects of Model 1. How much variation do we account for in student differences with those control variables? And then we compare that to the different models that have different control variables. There's a statistic that we're going to look at. Is there something we're looking for in that statistic? The answer is yes; there's actually a couple of things.

One, we want a high portion of variance. So if we had two models and two models only, and let's just say Model 1 accounted for 20% of the variance and Model 2 accounted for 60% of the variance, we would prefer the model that accounted for more variance relative to the one -- less variance. That's what we're looking for, a higher proportion of variance. But there's a point of diminishing returns. Suppose I now have three models. One of the models accounts for 20% of the variance, Model 2 accounts for 60% of the variance, and Model 3 has 25 covariates and it accounts for 62% of the variance, right? You've got a whole bunch of additional covariates that don't buy you much when you look at that proportion. They buy you 2% more. So there's no number that says is a 3% difference good enough, is a 5%? That's not what we're looking for here. We're not looking for a particular number; we're looking for a human judgment.

Do I really care? Is the difference between 60% and 62% enough that I would want to include all 25 covariates relative to including just 5? It's kind of what we're looking for here. So there's a point of diminishing returns. Why should we care about this? The model doesn't need to be needlessly complex. When you go out into the state and across the state and you're ambassadors for the model and people say, well, how do you control for differences between schools? You say, well, there are covariates for 1, 2, 3, and 4; and the teacher says why?

Now people want to control the model statistically looking at a single teacher's data.
going to show you is whether or not there are
different expectations for and using that as
criteria is your decision, right? You're going
to have the data by which you can make that
judgment. It doesn't matter in your view if --
we're going to show you the data.

Why don't I actually show you the data,
okay.

All right. The first statistic we're
looking at is for reading. This is the
R-Square. This is the amount of variation in
students' scores that the fixed effects account
for. Remember, refer back to your sheet so you
know which models are which because remember
some of the models include more covariates than
others, and we know that -- in fact, these two
models account for the largest proportion of
total variance. This is the one that has the
most covariates in it; this one has fewer.

Remember when I said there's a point of
diminishing -- in fact, they only differ in the
third decimal place. It's only because of the
way they're plotted that they appear to be
different there.

But look here; we see relatively similar
American Court Reporting
850.421.0058
1 judgment.
2 There will be differences in the estimates
3 of the teachers, but they are highly correlated.
4 MS. BOURN: If you look at the one with no
5 control variables, the two with no control
6 variables, 3A is virtually the same as the other
7 ones with all the variables, and the difference
8 between 3A and 3A1 is just one year or two
9 years, so isn't it the number of years that
10 seems to make the difference?
11 DR. DORAN: Ronda, you're a step ahead
12 because we're going to look at another criterion
13 in terms of the lags that tells us whether or
14 not, including more likely it doesn't add up or
15 not, but you are right. We're looking at
16 something that does seem to matter whether or
17 not it follows here or somewhere else on this
18 characteristic. But there's something else
19 that's different about these models, right?
20 That's why we wouldn't make judgments about the
21 models looking at any given criterion but only
22 looking across the different criterion.
23 Different lags, it does matter.
24 Now one of the things that's going on here
25 -- and this is the debate in the value-added
American Court Reporting
850.421.0058

237
1 literatures; do covariates matter at all? Do
2 you capture enough of the variability in prior
3 scores by conditioning on or by using pre-test
4 scores? Pre-test scores seem to capture a whole
5 lot of the variability in student scores because
6 remember that's what these models are doing.
7 They have the lags only, either one or two lags.
8 But when you have only one or two lags, they're
9 comparable when you have one or two lags plus a
10 whole lot of other things. So do those other
11 things buy you anything?
12 DR. COHEN: Harold, we should acknowledge
13 the point John made -- I think John made it
14 early in the conversation that while in the
15 aggregate these statistics, the teacher effects
16 tend to be correlated across the different
17 models like 0.9, 0.91, 0.92. They're very
18 highly correlated. But for an individual
19 teacher, they may differ. Say you have that one
20 kid who has terrible attendance in your class,
21 and if attendance matters then while it may not
22 improve the overall fit of the model in any
23 noticeable way, it may make it different for
24 some teachers.
25 MS. BROWN: This is what I'm taking from
American Court Reporting
850.421.0058

238
1 this, is that right now we're only looking at
2 the variance in the models, and because like you
3 said we're going to look across this whole
4 array, but this is telling us that it makes no
5 difference as far as accounting for the
6 variance; it may make a difference somewhere
7 else. Therefore, it's no harm, no foul
8 including or excluding when you're looking at
9 accounting for variance.
10 DR. DORAN: This is accounting for by fixed
11 effects. I'm talking about the control
12 variables. The control variables add a whole
13 lot more in terms of proportion of total
14 variance, but are there other possible
15 consequences? Yes. And remember, that's why
16 we're presenting along this series of the
17 different criteria.
18 MS. BROWN: So if, in fact, no harm/no foul
19 at the aggregate -- the big scale level -- then
20 and if there's the potential that at one teacher
21 level there might be a difference, it doesn't
22 hurt either way when we get to the final
23 decision with respect to controlling for the
24 amount of variance.
25 DR. DORAN: I just switched a moment ago to
American Court Reporting
850.421.0058

239
1 MAB (ph). You see a very similar thing here in
2 MAB. By similar, we don't see that this model
3 accounts for very little variability or these
4 ones don't account for little variability while
5 these models account for a whole lot of
6 variability. We see comparable estimates in
7 terms of how much variance in scores the
8 different models account for.
9 So part of the question that we're looking
10 at here is, in terms of accounting for variance
11 test scores, do you buy a lot when you add in
12 more covariates? Do we?
13 PANEL MEMBERS: No.
14 MR. LeTELLIER: Question about that. This
15 is looking at State data.
16 DR. DORAN: It's across the state.
17 MR. LeTELLIER: So as you're looking at
18 State data, obviously there's going to be less
19 variance because you have such a great number.
20 As you go down to the district level and then if
21 you went down to the school level and then down
22 to a grade level within the school, would there
23 be as you went down each step of the way and you
24 have less students that you were looking at,
25 would the variances on these be a lot greater?
American Court Reporting
850.421.0058
DR. DORAN: No idea.

DR. COHEN: I can answer part of that. To the extent you start to truncate the variance in student achievement, you're going to change the proportion of variance accounted for, but the models should hold pretty well through everything. All this is grade level specific; it's not a cross grade. So the grade -- district is going to look pretty much like the State. So while you might have small differences within the model, you wouldn't expect to see big differences.

MR. LETELLIER: Then as you finally went -- let's say you're using 7th grade, correct?

DR. COHEN: Yes.

MR. LETELLIER: So you're using 7th grade just in one school, say there's five 7th grade classrooms, and looking at just those five compared to each other.

DR. COHEN: You would -- when we say variance, the variance is explained by the control variables in the current score, in your test scores. Your FCAT score this year, right? Your most recent FCAT score.

If you were to go to, say, a trigonometry course in 7th grade where you have only the brightest students, I mean, that's a few years, then you would have very little variance in that dependent variable, so as a proportion this model would be explaining very little of that because there's very little variance there to explain. So it's not exactly -- it's not always the right question to ask, but when the best fitting lines don't fit the same, odds are you can probably also find where it was. Did that help?

MS. BROWN: I think what John's trying to say is, if this was all 7th grade Algebra 1, just say that, that way you're not changing levels of courses, you're not changing abilities; this is what it is. If this is the State level and we're saying that it accounts for approximately 70% of the variance, would it then hold true that let's say if we got to a district level or a school level for the same exact course, even though the level of variance might be different, but would they all be consistently the same? Is that your theory that you're talking about?

DR. COHEN: I'm not sure I -- see, the

American Court Reporting
850.421.0058

variance here, the total variance is the statewide variance of all students who were in any math class. So if you start truncating that variance by choosing, say, only Algebra 1 students in the 7th grade --

MS. BROWN: I know, but what I was saying was let's hypothesize that what we're looking at is Algebra 1. So we're not truncating, we're just saying; I'm just trying to do that as a very simplistic example. I mean, the point here is the models react similarly to the inclusion of the covariate in how they control for the variance in test scores, correct?

DR. COHEN: That's right.

MS. BROWN: Let me ask it another way. If you were to plot this graph 67 times one per district, would it look identical?

DR. DORAN: Okay. There's an answer to that question. This is on the statement. This is population. To the degree that districts are a representative sample of the State at large, they would look exactly the same, but they're not.

MS. BROWN: That's what I'm saying. They're not.

American Court Reporting
850.421.0058

class in 7th grade where you have only the brightest students, I mean, that's a few years, then you would have very little variance in that dependent variable, so as a proportion this model would be explaining very little of that because there's very little variance there to explain. So it's not exactly -- it's not always the right question to ask, but when the best fitting lines don't fit the same, odds are you can probably also find where it was. Did that help?

MS. BROWN: I think what John's trying to say is, if this was all 7th grade Algebra 1, just say that, that way you're not changing levels of courses, you're not changing abilities; this is what it is. If this is the State in terms of their student characteristics that districts systematically differ from the sample of the State, the model would hold and would look exactly like this. To the degree that the districts are a representative sample of the State, the model would hold and would look exactly like this. To the degree that districts systematically differ from the State in terms of their student characteristics in the population, it will be different. We cannot give you an answer in terms of would it be high or would it be low? It is unanswerable.

MS. BROWN: That's what I wanted you to say because that --

MR. MOREHOUSE: That's precisely the problem. Instead of a known impact on those teachers, they may end up losing their job. That number could be much more significant than we realize. I mean, it's one thing to try to achieve parsimony, but there's a human element that's involved here.
DR. DORAN: There is a human element. I'm going to go back to this. We should have shown this graph. Suppose we take a model that has no covariates and a model that has a whole lot of covariates and on the scatter plot, the correlation between those models was really close to zero then we would be able to say this matters a lot to teacher classifications, but it doesn't. We should show you that the correlation between the teacher effects under the different models is so highly correlated that it doesn't change those. It does some. Now, why, I don't know, it's going to change for a number of teachers, but in large part it does not change.

MR. LeTELLIER: Can you -- I know part of the thing is, you know, last time we asked you to run certain things and all that, and I have no idea how hard it is to run stuff, so I'm asking can you run something -- numbers for three different counties or two different counties that are completely different to see? Because one of the things is if you're saying the average county -- the average is here, but we could have counties, let's say that there's

American Court Reporting 850.421.0058

four or five counties that fall well below these averages, and that by not including the variables for those counties, for those schools, it would make a huge impact. I think in that case that's what we're looking at because we -- on the statewide level, fine, they all look the same, but --

DR. DORAN: I want Juan to weigh in on this in just a moment here. One of the things that I understand, this is a statewide model, but supposing we run this on different districts and we see differences. The models won't be run district by district. So I'm not sure it -- while it might be interesting to look at in terms of the policy, in terms of how this model becomes implemented and operationalized, what would the question be that would impact its operational status? So that would be my question.

MS. FEILD: No, but I was going to say you're right, but the issue is going back to Anna's comment or someone else, if we choose not to include the complex model that had 20 variables because we don't want to say to teachers, yeah, we included this, this, this, or

American Court Reporting 850.421.0058

better job in forming student predictions for all students in the state than this model does here. So the criteria by which we evaluate this is not in terms of its transparency to explain, but does it buy you anything statistically to do a better job in forming student predictions, and this model here, for example, doesn't do a substantially better job in forming those predictions, and this model versus a school that's very different. So to me it's an issue of going with a very simple model because it's easier to explain or with a complex model that will help us pick up all the differentiation from the diversity.

DR. DORAN: Now before I go over to you, just one second. Let's be clear. We're not recommending to you to choose one particular model over another because it's easier to explain.

MS. FEILD: No, no, I understand. I understand.

DR. DORAN: Whether these control variables do a better job in predicting where students should be and this statistic is showing whether or not including the variables does a better job in forming those predictions, and this model here, for example, doesn't do a substantially better job.
again very even distribution in terms of accountability of variance; then I think everybody would be comfortable in buying the argument that parsimony matters and we'd take the simple model. If we find out that there are significant variances across the districts, then if I took Anna's point there's reason to believe that accepting the more complicated model matters for some people, even though if you look at it at the State level, you can argue that it doesn't; individual districts you can argue that it does; and there

MS. BROWN: There have been truncates all the way down to individual teachers.

PANEL MEMBERS: (Over-speaking.)

MR. FOERSTER: Which is where Lawrence has been talking and John has been talking, so I guess the question -- I'm assuming the committee would like to see that if it's possible to do those kinds of calculations. Is it possible?

DR. COHEN: I would expect it's probably possible to do a comparison for two or three districts for overnight and look at the R-Square. Now if the variance in student achievement -- this is the R-Square -- if the variance in student achievement is different across the different counties that we look at, you will see differences and that's just a fact of life. The more you truncate the variance, the lower the proportion of variance explained is going to be. Also, so if we're going to do that, we'd like to do it with -- and if that's a statistic you want to look at, we should probably do with districts that have a lot of variation in student achievement paralleling the different State. You also get if you truncate the current score variance, like the FCAT variance by choosing say very low performing districts, you'll also change all of the co-efficients in the model, not because the world operates different there but from a statistical artifact. Let me just draw this real quickly.

MR. FOERSTER: Are you going the same place?

MS. BROWN: I don't know.

MR. COPA: Let me try something. Back to -- I think Harold mentioned it, we're developing a State formula, so we're not going to be calculating 67 different formulas, for example.

American Court Reporting 850.421.0058
They're huge, they're significant, they make a big difference. There's a bunch of indicator variables that we have to include for technical reasons, all kinds of stuff that goes in there for technical reasons that I can talk about if you want. These are the substantive variables that you guys wanted included and that will be looked at, and the yellow highlight tells you which ones were statistically significant. Remember, we had the variances explained across these three models were not very different, but you do see some things that show up as statistically significant and we can then walk through these and think about whether you want to keep them in the model.

So language impaired -- these are all SWD variables, all the different SWD variables. The more things you include, the fewer of them are statistically significant. That's what I was saying about introducing things that are correlated with one another. But in general many of the SWD variables are statistically significant. You might want to leave them in there. You may want to go through and say let's keep these and let's get rid of those, all the substantive variables that we have to include for technical reasons, all kinds of stuff that goes in there.

Homogeneity in classes. Oddly, in the first class it's not significant; in the second class it is, but it's a very small effect. That's probably where you deal with noise. These are things that appear only in the kitchen sink model we call it, Model 3C.
for your prior scores. So some of these things, even though it doesn't improve the fit of the model, it will make differences in expectations for individual teachers. So a teacher who's got a kid who's absent a lot, if two or three kids are absent a lot, the expectation even though the R-Square doesn't change, the expectation for what the kids will do and therefore the standards to which they're being held will vary a lot if you include the attendance.

MS. WOODHOUSE-YOUNG: You've highlighted some negative values and then I see up where you have negative pinpoint 0.8, the negative 8.85, that's not highlighted. But then we have highlighted here negative 7.0. I don't understand the number, just the values, why some are highlighted. I understand the negative and positive graphs maybe.

MR. FOERSTER: What's the range? What do the numbers mean?

MS. WOODHOUSE-YOUNG: I don't understand why some things are highlighted and why some aren't.

DR. COHEN: Okay, okay. If something's not highlighted, that means in these models we couldn't distinguish it from chance. It was not statistically significant.

MS. WOODHOUSE-YOUNG: So the numbers don't mean anything, though, the negative 10.08?

DR. COHEN: If it's white, probably ignore it because it could just be due to sampling error. If it's yellow, that means that it is not due to change.

MS. WOODHOUSE-YOUNG: So that negative 7.92 that's highlighted -- I can't see what it's actually related to -- and then the negative 5.36, that's highlighted?

DR. COHEN: Yeah.

MS. WOODHOUSE-YOUNG: So the numbers themselves, what does that mean to me?

DR. COHEN: These variables are coded as a 1 or a zero. That means that a kid who has been coded as other health impaired, his expected score, his expected growth is going to be almost eight points less than the kid who doesn't have that condition. If you put other things in the model, it's minus 5. These two numbers are probably not different than each other. That's within chance, but just due to the other things you're including in the model. So all the SWDs, you can take -- for all the SWD variables, you can take the difference and it's just a straight point difference in what you expect their score to be.

MS. MARSALA: How come SWD 7 goes from a negative to a positive?

DR. COHEN: SWD 7?

MS. MARSALA: Everything else stays the same.

DR. COHEN: This one is barely significant in a very large sample. Probably the other things that are highly correlated to this, I would bet that this is correlated with attendance, that students with emotional behavioral disorders are probably not attending.

I don't know that because I didn't look at the data.

MS. MARSALA: It's the next one.

DR. COHEN: Oh, there's --

MS. MARSALA: It's a negative 2.7, the positive is whatever they're expected to go higher.

MS. BROWN: On one model they're expected to go down and --

MS. MARSALA: Right.

American Court Reporting 850.421.0058

1 couldn't distinguish it from chance. It was not statistically significant.

2 statistically significant.

3 MS. WOODHOUSE-YOUNG: So the numbers don't mean anything, though, the negative 10.08?

4 DR. COHEN: If it's white, probably ignore it because it could just be due to sampling error. If it's yellow, that means that it is not due to change.

5 MS. WOODHOUSE-YOUNG: So that negative 7.92 that's highlighted -- I can't see what it's actually related to -- and then the negative 5.36, that's highlighted?

6 DR. COHEN: Yeah.

7 MS. WOODHOUSE-YOUNG: So the numbers themselves, what does that mean to me?

8 DR. COHEN: These variables are coded as a 1 or a zero. That means that a kid who has been coded as other health impaired, his expected score, his expected growth is going to be almost eight points less than the kid who doesn't have that condition. If you put other things in the model, it's minus 5. These two numbers are probably not different than each other. That's within chance, but just due to the other things you're including in the model. So all the SWDs, you can take -- for all the SWD variables, you can take the difference and it's just a straight point difference in what you expect their score to be.

9 MS. MARSALA: How come SWD 7 goes from a negative to a positive?

10 DR. COHEN: SWD 7?

11 MS. MARSALA: Everything else stays the same.

12 DR. COHEN: This one is barely significant in a very large sample. Probably the other things that are highly correlated to this, I would bet that this is correlated with attendance, that students with emotional behavioral disorders are probably not attending.

13 I don't know that because I didn't look at the data.

14 MS. MARSALA: It's the next one.

15 DR. COHEN: Oh, there's --

16 MS. MARSALA: It's a negative 2.7, the positive is whatever they're expected to go higher.

17 MS. BROWN: On one model they're expected to go down and --

18 MS. MARSALA: Right.

American Court Reporting 850.421.0058
1 That's not highlighted. So I think the point
2 system is not clear in my head anyway. I don't
3 know.
4 DR. HOVANETZ: Jon, just to be clear, these
5 are developmental scales, those points. So
6 think about it; we're talking about a
7 developmental scale for the FCAT which is 0 to
8 3,000; and not putting this necessarily in the
9 context of school level accountability, but I'm
10 putting it in the context of school level
11 accountability. When you're looking at a
12 student in reading going from grade 3 to grade
13 4, the expectations in reading is that they
14 learn 280-some points in order to make a year's
15 worth of progress. So when we're talking about
16 two points on the developmental scale score for
17 a specific learning disabled student, the swing
18 of four points on a 3,000 point scale or when we
19 talk about a year's worth of knowledge and a
20 year's worth of time for school accountability
21 purposes, the minimum expectation is 77 points
22 and that's 9 to 10. So the two points from the
23 policy perspective is not huge movement on that
24 developmental scale. So keep that kind of as
25 your context that, yes, they are specifically
American Court Reporting 850.421.0058

261 1 significant, but what does that mean on a 3,000
2 point scale.
3 DR. COHEN: Pam?
4 MS. STEWART: Just so I'm clear on this,
5 when we looked at, for instance, the other
6 health impaired that has a negative 5.36 on the
7 Model 3C, that would indicate that their
8 extensive DSS was 5 points lower than other
9 students in that same range of prior year FCAT
10 scores?
11 DR. COHEN: Yes.
12 MS. STEWART: Not just overall everybody,
13 but as you look at comparison with other --
14 DR. COHEN: Right, it compares kids by the
15 same prior year's score, the same ELL, the same
16 attendance. Everything else being equal, maybe
17 five points lower.
18 MS. STEWART: Right.
19 MR. TOMEI: Just a minor point. What was
20 your P-value calculated?
21 DR. COHEN: I don't know offhand. I can
22 get them for you. We took just a little over
23 the 0.05. We took two standard errors out and
24 highlighted it yellow. There were a couple
25 among the SWD things that were only marginal,
American Court Reporting 850.421.0058

262 1 but the rest are pretty decent significant
2 effects. Remember, we're dealing with a very
3 large sample here.
4 MR. TOMEI: But for some of those
5 individual categories, you may not be dealing
6 with a very large sample.
7 DR. COHEN: That's right, that's right.
8 For some of the individual categories you may
9 have very few kids, particularly the multiple
10 dual --
11 MR. TOMEI: Right, which is why you see
12 large numbers up there that aren't statistically
13 significant in a small --
14 DR. DORAN: But standard error --
15 MR. TOMEI: So my next question, we're
16 actually looking at anticipated variances in the
17 outcomes. Does that equate to an effect size
18 for all intents and purposes or would that be --
19 DR. DORAN: Yes, it's a natural effect,
20 it's an effect on the scale that you're
21 interested in seeing. So for example, I'd like
22 to talk about these things call effect sizes,
23 and an effect size is sort of a metric that we
24 can use to make a judgment on. That's what
25 Lance is talking about here.
American Court Reporting 850.421.0058

263 1 Here we have a natural effect size. We
2 don't need to convert it to anything because the
3 effect is a 13.7 difference, and so if we
4 converted it to a standardized effect it would
5 be interpretable to you and to Ronda, but here
6 it's a natural effect on the scale of
7 measurement.
8 MR. TOMEI: The reason I ask that question
9 is because of the earlier comment. If you're
10 looking at a 3,000 point scale and you see a
11 plus or minus two potential on two different
12 models for specific learning disabilities,
13 although it's statistically significant because
14 that's probably a large end population across
15 the entire state, when you're looking at an
16 effect size of about 2 on a scale of 3,000, you
17 have to wonder how useful is that regardless of
18 whether or not it's statistically significant.
19 DR. DORAN: There's practical significance
20 and there's statistical significance, and
21 they're not one in the same.
22 I am going to go to the next slide and get
23 you through the last piece of criteria. Is that
24 okay, Sam? Because there's a long conversation
25 that this group needs to have without me. The
American Court Reporting 850.421.0058
last thing that we're going to look at for now;
there's a lot more to look at -- there's a lot
more data to look at. We're not going to get
through it today. We're going to look at the
question of whether or not you should include
one or two lags or one or two prior test scores.
So what's the question? Should the
value-added model include one or two prior
achievement test scores for students? Remember,
when we say one or two we're talking about an
independent variable, so if we include two that
means we actually have three test scores: The
dependent variable, the current score, and then
the two prior scores.
So we're going to look at the standard
ersors again because those statistics are
relevant in helping inform this decision, and
what we're going to look at -- evidence in favor
of a desirable model -- is the same thing in
lower standard error so we can find anything in
terms of precision, and what do we care? Well,
as you bring more information into the
statistical model, you may or you may not do a
better job in forming a more precise teacher
effect. If you bring in more information, but
American Court Reporting
850.421.0058

it doesn't buy you anything in terms of
precision then we would ask the question why are
we doing it?
But if you bring in more information and it
buys you a lot in terms of precision then maybe
it's something that's reasonable to do. Well,
those are the judgments that you'll be looking
at here. This is the question, this is the
statistic and what we're looking for and why you
should care.
So here what I've done instead of taking
all of the models, I took the liberty of
choosing two models that were comparable but
different only in terms of the number of lags,
and 3A1 has the one and 3A has both. These are
boxed plots that we looked at at the beginning
of the day of the standard errors. In fact,
these are the exact same if you went back to the
box plot; and if you want to compare it you can
certainly do that, but for here looking at the
comparison, what we see here again that black
dot showing that we have smaller average
standard errors in the model with the two lags
relative to the model with one in reading, and
here's math; the differences are not as
American Court Reporting
850.421.0058

pronounced in this subject as they are in
reading, but the difference is there. We see
smaller average in standard errors in two lags
than we do with just one.
Now if you want to make a more holistic
judgment, you can turn back to the box plots
that we showed you of the standard errors and
you can look at all of them. So we're not
robbing you of that information. But for sake
of making a direct comparison, we choose two
comparable models, comparable, and they're three
evels and some other characteristics and they
differ only in terms of the lags.
So what other observations do we make here
that are meaningful? Anything?
This is a relatively straightforward
criterion.
MR. LeTELLIER: It just looks like there's
less error when you go two years.
DR. DORAN: Looks like particularly in
reading the estimates appear to be a bit more
precise relative to what you observed in math.
MS. FEILD: The problem is, though, you
have a lot of grade levels that by nature of the
grade level to go back to the reading, you're
American Court Reporting
850.421.0058
not going to have teachers having that much
data.
DR. DORAN: Now suppose this group were to
say we're going to include two lags. That would
not necessarily mean that you would eliminate
estimating value-added effects for fourth grade
teachers because there you'd have to have that
decision that you only use the one lag.
MS. FEILD: Right.
DR. DORAN: Now in terms of -- let's
explore that further. Suppose you're a fifth
grade teacher and every single kid in your class
has only one prior test score. You still
estimate the model, so I think the policy
decision he is not to always use two lags; it's
American Court Reporting
850.421.0058

Announce that we have smaller average in
standard errors in two lags than we do with just one.
Now if you want to make a more holistic
judgment, you can turn back to the box plots
that we showed you of the standard errors and
you can look at all of them. So we're not
robbing you of that information. But for sake
of making a direct comparison, we choose two
comparable models, comparable, and they're three
evels and some other characteristics and they
differ only in terms of the lags.
So what other observations do we make here
that are meaningful? Anything?
This is a relatively straightforward
criterion.
MR. LeTELLIER: It just looks like there's
less error when you go two years.
DR. DORAN: Looks like particularly in
reading the estimates appear to be a bit more
precise relative to what you observed in math.
MS. FEILD: The problem is, though, you
have a lot of grade levels that by nature of the
grade level to go back to the reading, you're
American Court Reporting
850.421.0058
not going to have teachers having that much
data.
DR. DORAN: Now suppose this group were to
say we're going to include two lags. That would
not necessarily mean that you would eliminate
estimating value-added effects for fourth grade
teachers because there you'd have to have that
decision that you only use the one lag.
MS. FEILD: Right.
DR. DORAN: Now in terms of -- let's
explore that further. Suppose you're a fifth
grade teacher and every single kid in your class
has only one prior test score. You still
estimate the model, so I think the policy
decision he is not to always use two lags; it's
American Court Reporting
850.421.0058
only have we're only going to include one, we have to say, well, by the way, that means your error of measurement is going to be bigger, your standard error.

DR. DORAN: So remember this is one factor that plays into the standard error measurement. Teachers in grades 5 and up would have the benefit of having possibly, possibly smaller standard errors because we're using more information, but it's not a guarantee.

Teachers in grade 4 can still have small standard errors because there are many factors that are used in terms of creating the standard of error, not only the lag. But they would have the down side of not being that extra information, so that would be something that would -- it's just an artifact that you don't teach testing second grade.

MS. FEILD: But that's compounded by what model you choose as to what covariates because if you choose not to use any covariate at all, which would be Model 3A, right, then really the lag of two years versus one is the biggest impact on your score; am I correct?

DR. DORAN: I don't remember exactly, but American Court Reporting 850.421.0058

---

we're in good shape there. I also want to say that it is a distinct honor to be a part of this group. I mean, I am really astonished at the quality of the discussion that has taken place already today and I hope that you guys feel equally gratified.

I have every confidence that we have lots of people around the table that see this. The struggle is getting us all to see what one another sees and that has proven to be challenging.

Where I think I would like us to go is this: An effective strategy last time was ruling things out so that we can focus on the things that we want to keep in play. That having been said, I don't want to move us down the path any more quickly than you guys are comfortable with. So if you're uncomfortable with the rate at which we're marching down this path, please anybody jump in and say I'd like to talk this one over some more before we put things to a vote and scratch things off the list.

That having been said I'm going to throw out where I think the temperature of the room is American Court Reporting 850.421.0058
<table>
<thead>
<tr>
<th>Page</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>272</td>
<td>1 with respect to some of these models so that we can get a sense of where to start. Is it fair to say -- I'll start with the easiest one first -- Model 4, the sustained differences model that has the lowest amount of precision, and the least number of variables folded into it. This doesn't appear to be where any of us wants to go. Is that a fair assumption? So could we put that one to a vote that the committee would like to cease consideration of Model 4, the sustained differences model. MS. FEILD: So moved. MR. LeTELLIER: Second. MR. FOERSTER: I love it. Thank you. All in favor? DR. HOVANETZ: Remember hold your hand up. MR. FOERSTER: Yeah, we've got to get the camera around. Okay. Thank you. We'll scratch that one off the list. Where can we go next? The one lag models. Are we all comfortable that we want to put in two lags and do we all understand that we're not saying that we're going to include only data for American Court Reporting 850.421.0058.</td>
</tr>
<tr>
<td>273</td>
<td>1 which we have two scores, but when we have that data we're going to use two scores? When we have only one, we'll use it. So that would mean that what we would -- MS. FEILD: So moved. MR. FOERSTER: Yeah, don't &quot;so moved&quot; me. Give me a motion. Which ones have one lags? PANEL MEMBERS: One and 3A1. MR. FOERSTER: Okay. So the motion is that we will cease consideration of Model 1 and Model 3A1, is that right? MS. FEILD: Yes. MR. FOERSTER: Second. MR. TOMEI: Second. MR. FOERSTER: All in favor? Oh, this is fun. Thank you very much. Okay. That leaves us with Model 1A which does not include school effect and variance of Model 3 which do include school effect. I think where we're at after much discussion on school effect, and I don't think we're done with that discussion, but we all agree that it matters and we do want it to be taken into account somehow. The how is what remains to be determined. Is that where everybody's at? American Court Reporting 850.421.0058.</td>
</tr>
<tr>
<td>274</td>
<td>1 Okay. So can I have a motion that we'll cease to consider Model 1A? MR. TOMEI: So moved. MR. LeTELLIER: Second. MR. FOERSTER: All in favor, raise their hand? Okay. Thank you very much. That leaves us with the three variants of Model 3, one of which we include no additional covariates, one in which we include just the basics which would be ELL, SWD, gifted, and attendance, and then the kitchen sink variety. I think again given the discussion that we've had to this point that we're all in favor of including additional covariates, maybe lots of them, which would mean that Model 3A is not something we want to consider any further. Is that where we're at? PANEL MEMBERS: Yes. MR. FOERSTER: So can I have a motion that we -- the committee wishes to cease consideration of Model 3A? MS. BROWN: So moved. MR. LeTELLIER: Second. MR. FOERSTER: All in favor? Thanks. Okay. We're honing in, I think. Most of us feel like we want to factor in some of the additional covariates beyond ELL, SWD, gifted, and attendance, and we need to discuss which ones and how and why and what the implications of that are, but we don't really want to be considering some aspect of the kitchen sink model, which is 3C. Is that a fair statement? PANEL MEMBERS: Yes. MR. FOERSTER: Okay, then I need a motion that we wish to cease considering Model 3B. MS. NOYA: So moved. MR. LeTELLIER: Second. MR. FOERSTER: All in favor? Thank you. Okay, time to go home. Model 3C. We stopped when we were looking at the list and I'm hoping we can get the slide back up so everybody can look at it. The list of covariates, some of which were found to be statistically significant, some of which were found to be statistically not significant; and I think we can pick up discussion with which of those covariates we want to be included; and I'm going to do my best here to talk through some of the factors that should be taken into consideration when we're talking about that. American Court Reporting 850.421.0058.</td>
</tr>
</tbody>
</table>
With respect to the factors that are statistically insignificant, that means that they do not help in a predictive capacity at all, and we know that for sure because we looked at all the data district-wide, and there's no evidence to suggest that incorporating those things give us any ability to predict student outcomes any more accurately than not including them. That having been said, it's hard for me to imagine a scenario where we would to be talking to people about why those things are still in there because AIR has done the work. We can say for sure it'll matter. That's my opinion.

The counter-point could be that keeping them in does no harm and it gives us the opportunity to explain to teachers who might be impacted by one of these categories -- say, hearing impaired, visually impaired, emotionally, behavioral, these factors that do not have statistical significance -- it may be politically useful to say that those have been left in the model.

DR. COHEN: I may have left a slightly wrong impression. This is a general pattern, MR. FOERSTER: That's an important point. DR. COHEN: I mean, it's not all that difficult but yes, this is a -- PANEL MEMBERS: (Laughing, talking, over-speaking.)

but remember we estimated 112 different models or something like that. In some of them, some of the particularly SWD variables, some of them pop up as significant in other grades or other subjects. I think that's what it's based on, not a grade 7 math.

MR. FOERSTER: That's an important point.
DR. COHEN: I mean, it's not all that difficult but yes, this is a --

PANEL MEMBERS: (Laughing, talking, over-speaking.)

DR. COHEN: Sam.
MR. FOERSTER: Yes, Jon?
MR. LeTELLIER: With the knowledge that we just had, let's look at number SWD 12, traumatic brain injury, and let's say that some of those other grades -- that there was a significant number. How would we if we decide to take out something like that, how would we explain that we're going to allow for the other ones, but if your son just had a traumatic brain injury that it's not statistically significant?

MS. BROWN: Or to the teacher. I move that we include them all.

PANEL MEMBERS: Second.

We factor these in, particularly if we keep in things that have proven to be statistically insignificant?

DR. COHEN: But we only have that for this grade. That's the problem, what you just said.
MR. TOMEI: We can fix that, though, Jon.
DR. COHEN: We have all the data here, we still don't have it for each and every grade.

Let me just make the case of a dual sensory impaired, just as an example. The most likely value for that typical value of the population is that number. So the chance is greater than about a 5% chance that could be due to chance. Let's look at what statistical significance means. It is still more likely than not that that is an effect, that there's a positive --

I'm sorry -- a negative effect there.

So I don't even know how many dual sensory impaired students have it in the state. Did anybody say that?

PANEL MEMBER: Very small.
MR. TOMEI: Very minimal.

MS. WESTPHAL: But there's a potential that the reason if I'm understand this why it's not significant is because you only pooled --
MS. BROWN: But in a lot of those instances like emotional behavior -- I just picked one. Okay, let's take emotional behavior now. It's not showing a significant -- it's not yellow, but there's a chance that there's a teacher that has six kids in her classroom and all six of those kids are EBD.

DR. COHEN: Even if it is the best estimate of how much impact it has is only a point or two on the scale; within 7th grade the scale ranges hundreds of points. I think the typical growth in the 8th grade is on the order of 250 points, not the exact number but that's the right magnitude. On average, there's about a 1 or 2 point difference among these kids.

MS. WESTPHAL: I'm guessing because you pulled math we would see different numbers for reading.

DR. COHEN: I can -- hold that --

MS. WESTPHAL: But my point is, okay, let's just take that out of it and say there is one that's more significant, dual sensory impaired. Maybe those kids are concentrated at the school for the deaf, for the blind, for example. There

American Court Reporting
850.421.0058

could be a teacher who only has those students, so let's say she has four students in her class. All four of her students are dual sensory and for her or him it's going to make a big difference if we don't. Otherwise, would we not be throwing them into the general population?

DR. COHEN: Otherwise we would be throwing them into the general population. But that is not statistically significant. Really, what that means is that it says we're not 95% certain that this is different than zero, right? But in the data we have in this sample, the average score is 121 points less than you would expect of a very similar student who didn't have the same disability.

MS. WESTPHAL: So worst case scenario for that teacher, her scores come back and the statistician says, you know what? There's not enough data; we don't have a big enough population in your room to say whether you're highly effective, not effective; so we've got to put you right here and now your evaluation is going to take over the bigger percentage piece.

MS. ACOSTA: The business rule can control whether or not that particular piece of data is included if we want to just include everything, and my understanding from Harold is it 94% or was it 55% for some of these variables, but the data exists. So there's more work to be done to figure out how do we factor these variables properly and effectively and appropriately into the model if we keep them all? So I thought we should have that discussion before we --

MR. FOERSTER: I am so glad that you pulled the reins. I think it's fair to say that we have already narrowed things down a lot. I mean, we're down to one model and we're talking about which variables do we want to include and not include, and my understanding from Harold

American Court Reporting
850.421.0058

and Jon and Christy was that was essentially what we needed to try to get to happen this afternoon. I'm wondering if what we do instead of trying to nail this today since it is late in the day, I think we're tired, and I think we would all benefit perhaps with some reflection and some time to think tonight.

What do you guys think about handing it back to AIR and let's keep working through the agenda that they have prescribed for us, and we will take this issue up tomorrow as we put a finer point on what exactly we'd like to see in the model?

MR. LeTELLIER: If we do that, I think that's a great idea. Two things, one can we have them do some of that data that we were looking at --

MR. FOERSTER: For other grades and subjects?

MR. LeTELLIER: Yeah, and then the other thing is what Lance was saying; are we able to data-wise statistically make it so that we can include if we want to just include everything, include it and come up with a way to have the data be useful in a model or is it going to from

American Court Reporting
850.421.0058

used.

MS. WESTPHAL: If we don't put that in there then her students are thrown into the general population and she is going to look like she's not as effective. -- am I -- getting that?

MR. TOMEI: I just want to say that actually I'm in favor of keeping all the variables in, but I think we need to be cautious. What we're not looking at here -- we know the ones that are not significant were less than 95% certain, but what we don't know is, was it 94% or was it 55% for some of these variables, but the data exists. So there's more work to be done to figure out how do we factor these variables properly and effectively and appropriately into the model if we keep them all? So I thought we should have that discussion before we --

MR. FOERSTER: I am so glad that you pulled the reins. I think it's fair to say that we have already narrowed things down a lot. I mean, we're down to one model and we're talking about which variables do we want to include and not include, and my understanding from Harold

American Court Reporting
850.421.0058

and Jon and Christy was that was essentially what we needed to try to get to happen this afternoon. I'm wondering if what we do instead of trying to nail this today since it is late in the day, I think we're tired, and I think we would all benefit perhaps with some reflection and some time to think tonight.

What do you guys think about handing it back to AIR and let's keep working through the agenda that they have prescribed for us, and we will take this issue up tomorrow as we put a finer point on what exactly we'd like to see in the model?

MR. LeTELLIER: If we do that, I think that's a great idea. Two things, one can we have them do some of that data that we were looking at --

MR. FOERSTER: For other grades and subjects?

MR. LeTELLIER: Yeah, and then the other thing is what Lance was saying; are we able to data-wise statistically make it so that we can include if we want to just include everything, include it and come up with a way to have the data be useful in a model or is it going to from

American Court Reporting
850.421.0058

and Jon and Christy was that was essentially what we needed to try to get to happen this afternoon. I'm wondering if what we do instead of trying to nail this today since it is late in the day, I think we're tired, and I think we would all benefit perhaps with some reflection and some time to think tonight.

What do you guys think about handing it back to AIR and let's keep working through the agenda that they have prescribed for us, and we will take this issue up tomorrow as we put a finer point on what exactly we'd like to see in the model?

MR. LeTELLIER: If we do that, I think that's a great idea. Two things, one can we have them do some of that data that we were looking at --

MR. FOERSTER: For other grades and subjects?
<table>
<thead>
<tr>
<th>284</th>
<th>285</th>
<th>286</th>
<th>287</th>
</tr>
</thead>
<tbody>
<tr>
<td>what you're saying, are there some things if we included it no matter what we did, it would negatively affect things? Because just for myself as I think about it tonight, I just want to know what to kind of have ruminate around in my head in thinking.</td>
<td>think about it as kind of SWD is a whole, class size is a whole, modal age is a whole, mobility is a whole. To me there --</td>
<td>thought it would be a headache to pick and choose among them. Some of the things like class homogeneity, which is significant in one of the models; it's significant here and there, but it's a tiny effect. You need a class that had -- if you had a class that had a 100-point difference between the 25th percentile and the 75th percentile, you would have a 1 point difference. If you had a four point difference, you need basically two years of growth within one class. A very diverse class and that would still only count for four points.</td>
<td>attributed to. If most of the kids who have specific learning disability are in very homogenous classes, they tend to be other kids who are challenging in those classes, the exact effect of either one of those variables is going to be trading off. Sometimes one will be bigger, sometimes the other will be bigger, especially since there's not enough information in the data. So those correlations are the only things that are set for now.</td>
</tr>
<tr>
<td>MS. FEILD: I also want to know as I'm looking at these things, it's not 20 indicators really. I'm looking at them under categories, and are we talking about fine-picking and saying that we're going to go in and in terms of the SWD we're not going to include the dual sensory and we're not going to include visual, but we'll include the others. I mean, are we even going to get to that granularity? Do we want to do that or do we want to look at it as an overall; if this child is SWD, some of them maybe, you know, have positive/negative depending by different grade levels, so should we be thinking about it as a whole or are we going to sit here and say, okay, well, the dual sensory in grade 3, 5, and 9, it was -- it showed an effect, but in the other grades it didn't, so I'm not sure we want to do that. I just want us to think about that because I would like to keep all of the SWD variables as a group. Some of the things like class homogeneity, which is significant in one of the models; it's significant here and there, but it's a tiny effect. You need a class that had -- if you had a class that had a 100-point difference between the 25th percentile and the 75th percentile, you would have a 1 point difference. If you had a four point difference, you need basically two years of growth within one class. A very diverse class and that would still only count for four points.</td>
<td>There's a trade-off in what the effect is for four points.</td>
<td></td>
<td>23 of 38 sheets</td>
</tr>
<tr>
<td>Page 288</td>
<td>Page 289</td>
<td>Page 290</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>1 day.</td>
<td>1 you may or may not want to include that in the model. The problematic thing being that you have just established different expectations for kids. While that's helpful from the teacher evaluation standpoint and seems to level the field, it's problematic in that you have different expectations for kids. I just want to bring that up because it was a point of lots of conversation last time and I think we should be making these decisions with that in mind. Any thoughts on that?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS. KEARSCHNER: I don't remember, what's the difference between class 1 and class 2 and class 3 through 6?</td>
<td>MR. FOERSTER: Perfect. Thank you, Jon. Are there any other specifics that you guys would like to see prepared to inform tomorrow's discussion about how we're going to finalize a recommendation?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR. DORAN: This is the number of classes the student was enrolled in for the same courses. What's the definition --</td>
<td>DR. COHEN: I'm working on a little spreadsheet that shows some examples of that; I should have that in an hour.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS. KEARSCHNER: Subjects.</td>
<td>MR. FOERSTER: Sam, it's actually where we're going to next. We're going to show the consequences on the different expectations for different groups of kids, not for every single one of these particular categories. That's actually where we're going with the data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR. DORAN: The same subjects? There are some students who are associated with multiple classes for the same subject.</td>
<td>MR. FOERSTER: Okay. Before I hand it back over to you, committee, AIR is going to have one night to do some additional materials preparation, analysis, whatever. Can you or do you have any specific requests that you would like to see prepared for tomorrow? Ms. Bourn?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR. LeTELLIER: Could you put a slide up tomorrow because we're not going to vote on this today and we would be able to think about that? That might in parentheses just have those little things so that as we're looking at it, that would be easier, I think, for us to say that's based on this, that's based on this.</td>
<td>MS. BROWN: Did we ask AIR to do certain districts? Is that already on the agenda to be done?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR. HOVANETZ: It goes back to the finding of the variable that we did this morning and how we defined it, whether it was a cognitive variable saying if this student has this characteristic their expectation is this much.</td>
<td>MR. FOERSTER: Well, I think Juan made a huge hour-plus long discussion, and if I'm American Court Reporting 850.421.0058</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Court Reporting 850.421.0058</td>
<td>different versus the continuous variables that we talked about continuous variables being homogeneity, age, attendance where each incremental unit is associated with each incremental DSS point difference. So an increase in one day of attendance equates to an increase of an expectation of a 0.16.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR. TOMEI: I get that for the attendance points now. The other ones?</td>
<td>MR. TOMEI: What's the implication?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DR. HOVANETZ: So for mobility for each additional school transition, the expectation for their growth is decreased by five points. So the continuous variables is each increment the variable moves, the DS doesn't impact it by the amount that you see up there or the dichotomous variable which is basically a majority of the variable that we talked about; it's just one single expectation, if the student has this characteristic or trait, their expectation is older by that many.</td>
<td>MR. FOERSTER: In a variety of scenarios, I'm assuming, with one extreme being there is no school effect considered and the other extreme school effect is completely attributed to the school and perhaps some --</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR. FOERSTER: I guess we should point out, also, that there is this policy implication that we've talked a good bit about last time. Just because you see that a characteristic can be argued to weigh in on expected student growth, American Court Reporting 850.421.0058</td>
<td>DR. COHEN: I'm working on a little spreadsheet that shows some examples of that; I should have that in an hour.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
great point there and I'm going to do my best to explain why that was problematic.

PANEL MEMBERS: (Over-speaking.)

DR. COHEN: What I can do -- I have all the grade 7 data, for example, on my laptop and it's very easy to run the average teacher effect in each subject by district, and you guys can look through your own districts and say I like this model, I don't like that model because it made me look bad. Oh, doesn't it.

MR. LeTELLIER: I think there is some use to that and you've got to realize not all 67 counties are represented here, so it's not just for the county. It's just so much as what some of us were discussing, are there any variables that in some county might have heavier weight than another, that in the average across the state --

DR. COHEN: No, that we can do right now is run 67 different --

MR. LeTELLIER: No, no, I'm not saying to run 67 counties. Their whole thing was whether or not there might be a county that it will be impacted upon more, and remember we had discussed running two or three counties just to American Court Reporting 850.421.0058

see.

DR. COHEN: We talked about that, but let's say we find there is a strong negative co-efficient for students with emotional behavioral disorders in one county. Everyone else is zero or positive, this one county is negative. Do we then hold that county to a different standard and say, okay, we expect lower growth from you?

MR. LeTELLIER: No, it's just we're including as a variable across the -- we're including as a variable across the state and what we had talked about earlier unless I misunderstand this is that if there -- there may be in some instances -- maybe ELL is a good example where a specific county, it would impact them more. If the other counties, it doesn't matter if it you put it in or don't, it's negligible. But for two or three counties it's huge and it's real and it's statistically real for those counties. That's what we're --

MS. BOURN: Then we know that for that model we should keep it in, in order to have fairness across the board.

MR. LeTELLIER: Exactly, yes. Not that American Court Reporting 850.421.0058
to weight it 10%, 40%? That may differ by district, right, and including the variables you're talking about. If there is differentiation by grade, do we want to throw out SWD across the Board? Maybe not, I don't know.

MS. KEARSCHNER: Do we want to look at SWD in grades other than 7 math, which is what this is?

MS. FEILD: That's what I say. What do we need to look at in 3C particular to make the final recommendation? As for example, if we already include school effects which 3C does have, how do we want to weight it? Do we want American Court Reporting 850.421.0058
that we're creating different expectations.

MR. TOMEI: No, no, no.

MR. COPA: Let me just add one thing that might be helpful. Since the committee has narrowed it down basically to one model, I mean, we went through a whole bunch of slides. They basically estimated 112 models and they were just presenting grade 7 math and reading as an example just based on space. But now since we're down to one model, AIR can share all 14 grade and subject combinations for that one model so you can see the results for 4th grade, 5th grade, reading, math, et cetera.

MR. TOMEI: It might also be helpful, too, rather than us trying to amalgamate what we're looking at across 14 models, if we could see perhaps a list of any of these variables that were bound to be not significant for either subject in any grade level and what the greatest effect size was for those that were insignificant across all grades in both subjects, is that doable?

DR. COHEN: It probably is, but I'm going to need to write that down. I'm going to fill up a notepad to write that down. Okay.

American Court Reporting 850.421.0058

MS. FEILD: So if we're not comfortable, what happens?

DR. COHEN: Then we circle back and Sam has a hard job. We have to go back --

MR. FOERSTER: Yeah, this isn't in concrete but it seems like everybody was pretty -- we were going to 3C. I mean, we were heading that way, so maybe we keep marching down that path and see problems we'll back up.

DR. DORAN: So sort of big picture of where we are, you know, is we spent a tremendous amount of time this morning evaluating the models against some criteria, and you've come to at least what's a tentative conclusion about which of the models you favor more than others. But now what we want to do is start showing you some of the impact data. What's the impact on these model decision on expectations? What's the correlation of these with teacher with characteristics or student characteristics and so forth? Now you can make decisions about I even like this model more or now I have concerns about this decision, and I want to come back and revisit some of those issues.

American Court Reporting 850.421.0058

MR. TOMEI: I'm just wondering if we have any variables that prove to be insignificant for both subjects in all grades and then what was the greatest effect size for those, any one, just the single greatest effect size because that may tell us if there are any variables that really might just be worth not putting in the model that summary.

MS. FEILD: So my question is if we picked the model we have to decide if we want to include some of the covariates, right? What other decisions is AIR going to need from us by the end of the day tomorrow?

PANEL MEMBER: Percent.

MS. FEILD: A percent of what school effect, but what else because I'm not sure that there's other pieces that we haven't even discussed --

DR. COHEN: We want to show you some impact data. What does this model choice say about expectations for students with different characteristics, and which groups of teachers seem to do better or worse under this model, so that you can take a look at the impact of your decision and make sure you're comfortable with

American Court Reporting 850.421.0058
We've done a lot today. We've made some very important decisions. I leave it to the committee. We can keep going till 5:00 or we can stop now and reconvene tomorrow morning.

MS. BROWN: Can I ask a question? What's on the agenda for tomorrow because I need to know what we're adding to tomorrow to make sure so I can make that connection.

DR. HOVANETZ: The only thing, if you want to flip through the power point presentation is just the impact of the variable that we're talking about right now. We're picking up on the variable discussion and sharing more information. We'll have you fill out an evaluation before you leave, you can write down your specific questions that you had just like we did last time, we'll review those tonight, and we'll start in the morning responding to all of the questions that you all are leaving here with today. We'll do a recap of this day's discussion, so we'll just spend the first hour recapping and answer questions; and then literally just bagging it right back into where we are. So the stimulation on the school effect and how that impacted individual teachers under the model.

So we'll take 15 minutes. You can complete the evaluation, things that are still burning questions you'll be able to jot down. We'll answer those tonight and start the presentation.

given an opportunity to think about this, the consequences; you were given all of the information so that to the degree that we can do stuff tonight reliably and efficiently, we want to make sure that you have that information. That was sort of the goal of where we wanted to be today. So if you need anything, ask. We've got computers.

Let's look at a couple of other things. We're going to talk about the expectations. Recall we talked a little bit earlier about what these growth expectations are and I shared with you earlier that we're going to talk about -- go ahead.

DR. HOVANETZ: We're actually going to see if we can have Jon present the school effects conversation model, and this is a big, huge, weighty discussion that might be best served for us to start tomorrow rather than after 4:00 today, just because you've made a lot of progress and decisions and have a lot to chew on, we want to show the school effect impact stuff and then --

MR. COPA: Let me offer option CA 2, 3, 4.
tomorrow just recapping what we talked about
today and answer any unresolved questions that
you have.

MR. FOERSTER: Okay. Are you all
comfortable with Christy's plan?
MS. NOYA: Yes, I am.
DR. HOVANETZ: Okay. Don't go anywhere.

* * * * * *

(Whereupon, this concludes Day 1 of the
meeting.)

C E R T I F I C A T E

THE STATE OF FLORIDA )
COUNTY OF WAKULLA )

I, Suzette A. Bragg, Court Reporter and
Notary Public, State of Florida at Large,
DO HEREBY CERTIFY that the above-entitled
and numbered cause was heard as herein above set
out; that I was authorized to and did transcribe the
proceedings of said matter, and that the foregoing
and annexed pages, numbered 1 through 304,
inclusive, comprise a true and correct transcription
of the proceedings in said cause.
I FURTHER CERTIFY that I am not related to
or employed by any of the parties or their counsel,
nor have I any financial interest in the outcome of
this action.
IN WITNESS WHEREOF, I have hereunto
subscribed my name and affixed my seal, this 13th
day of June, 2011.

__________________________________
SUZETTE A. BRAGG, Notary Public
State of Florida at Large
My Commission Expires: 2/21/2013