Virginia M. Barry, Ph.D.
Commissioner of Education
Tel. 603-271-3144

Paul K. Leather
Deputy Commissioner
Tel. 603-271-3801

STATE OF NEW HAMPSHIRE
DEPARTMENT OF EDUCATION
101 Pleasant Street
Concord, N.H. 03301
FAX 603-271-1953
Citizens Services Line 1-800-339-9900

New Hampshire Task Force on EFFECTIVE TEACHING PHASE II

◆ MINUTES ◆

12th Meeting for Phase II

June 19, 2012

Londergan Hall ◆ Room 15

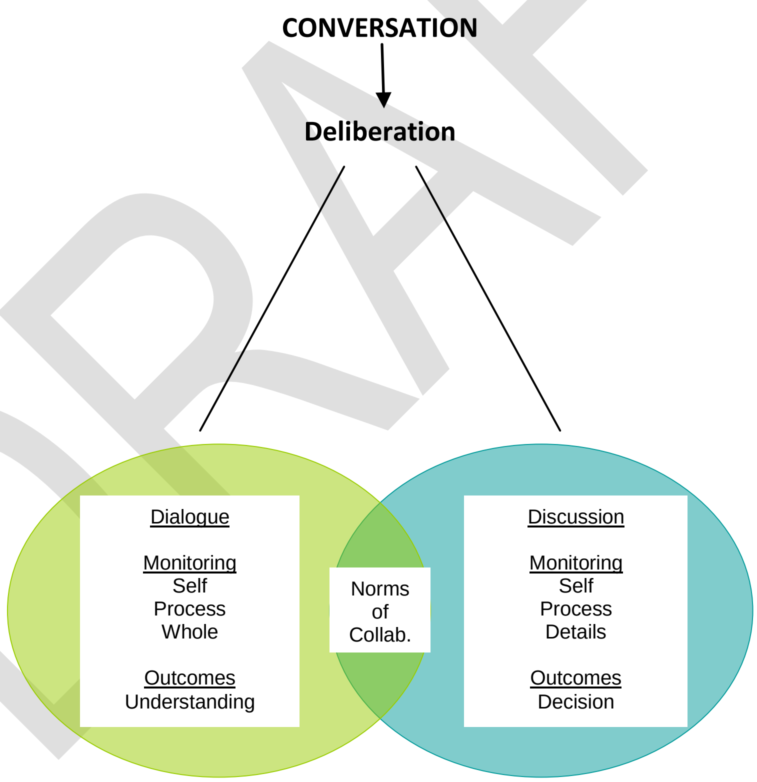
Commissioner Barry

The meeting began at 9:12 am with Commissioner Barry welcoming everyone. She then let everyone know that independent of their teachers and other educators, Maine decided to put laws in place that were approved. These laws are in relation to educator evaluations. She said that today's meeting was to tie things up, keeping in mind that next year is a pilot. The model will be presented to the superintendents on June 26, 2012. We have come up with a model that is flexible, supportive and comprehensive. Induction and mentoring needs to be in the schools in the next couple of years. Professional development must be tied to teacher evaluations. We are working very closely with our professional development centers and we are looking at what role the professional development centers will have and the quality they will offer our teachers. We are looking at the whole educational process in New Hampshire. She said that it is critical today to let some of the things go that people were worried about and trust that the system that we come up with will be introduced, it will be monitored, it will be piloted, but it will be a system that will come from educators and will not be imposed on us. We are ahead of other states and this is our opportunity. There was a meeting yesterday regarding the waiver and the biggest thing holding us back on the waiver was the educator effectiveness piece because the superintendents were saying if we come up with a system that the teachers and principals cannot buy into we do not want to move forward with the waiver. Now they are convinced we need to move forward with the waiver because it changes our ability to look at AYP differently.

Commissioner Barry then shared that she has rethought some things. If the group wants the 20% for teacher evaluation removed, she is prepared to support that. She stated that we need to look at student achievement as the primary goal of what we do as educators, but that we have multiple measures that districts can determine and we have to use the statewide test. Today is about staying focused on student achievement and coming up with a process. Nationally the recommendation is a minimum of two written observations per year. If you can consider that, it is important to put it in. Beyond that, once you set the standards, which you already have, this is the model that school districts will be expected to align with. They can fully adopt it or they can align it with our model. Another piece that has not been worked out is reporting. People are worried about how all of this will get reported. This is a serious decision that will not be made today. It has never been the goal to report on individual teachers. Commissioner Barry said that there has also been some concern about the resources. It is up to all of us identify where the resources will be and provide technical assistance when districts need it. The transition is five years. The first year will be a pilot year. The second year will be used to identify the areas you want to change. In the third year you need to show that your system is where you want it. And by the fifth year, we have adopted the system in total. We need to leave here today with a model so that it can be presented to the superintendents on June 26, 2012.

Kathy Dunne

Kathy gave an update on the SCEE team. They went to a conference out in Seattle and Irv Richardson was presenting. She then asked everyone to open their folders while so they could see what the contents were. The purpose of the day is to make decisions. There was a diagram that she presented to the group to show how to move a conversation into a decision.



She then made a chart of the decisions that had already been made and those that still needed to be made. The decisions that had been made were based on the Gradients of Agreement that we did at the meeting on April 10, 2012. She explained that for the most part everyone was in agreement on the decisions that had been made. There were a few varying opinions, but that is why they are gradients, they are not yes and no answers. Below is what it looked like.

DECISIONS

MADE	**TO BE MADE**
• Key Principles • Dimensions/Domains • Standards • General Evaluation Framework	• Specific Measurement Framework → Performance level names, descriptors → Evaluation frequency → Consequences → Use of state test data/SLO's → District requirements

Carol Kierstead

Carol said that we have already made most of the decisions that needed to be made. She then said that the charge for the day is to focus on the Specific Measurement Framework. There was a sheet titled "Decisions for the Task Force" in the folders that outlined each of the pieces that the group needed to make decisions on.

Scott Marion

Scott reiterated what Commissioner Barry said in that more information would be needed to present to the superintendents on June 26, 2012. He acknowledged that the model is still a work in progress, but they need something to present. He then said that after the superintendents meeting, they would come back to the task force and be looking for all or a subset of volunteers to read the model for coherence and logic and then start thinking about the next steps for implementation. Beyond that, there are a lot of things to decide about piloting it, reviewing it and implementing it.

Commissioner Barry then said that once we have agreement on the specifics that are being talked about today, the actual procedural elements, the superintendents are in agreement that they will align with the state model (standard pieces of it and the non-negotiable pieces) because of the flexibility. Some of the non-negotiable items are a minimum of two written observations in a year that are tied to student achievement. That is being determined by the district, including statewide testing, as well as the locally determined multiple forms of assessment that they are using. Once we get that agreement, we can move forward into the policy stage. She said that she is hoping Senator Stiles will be an advocate for us in this process. She said she is imagining two groups. One group will firm up the final procedures (that will be a written document that would come from the Department). The second group will be working with our policy makers to support the process, as well as the State Board of Education because they will also be involved in the process. Along with that, the higher education group is interested in continuing to support these kinds of efforts.

Scott added that the deadline for the waiver application is September 6, 2012.

Commissioner Barry

Commissioner Barry asked if everyone understood the waiver and what it represents. She said that the opportunity for the waiver is to eliminate AYP as we currently know it and to move us into a new generation of leaving NCLB behind and looking at how, as a state, we have developed systems that support quality education in the closing of gaps. The areas of the waiver are how we are addressing the Common Core, College and Career Readiness, data assessment using a systemic approach to looking at data in a coherent manner and educator effectiveness and focus on our priority schools. We are doing all of these things and we are in a really good place. The goal is to submit the waiver in September and within three months, (we have already frozen our AMO's for two years), this would put us into a whole new way of thinking about assessment. We would no longer have to report our lowest performing schools in the manner that we are right now. Other schools would not even be reported. And we would be able to move forward with our network approach.

Kathy Dunne

Kathy asked everyone to read Scott's latest iteration of "The New Hampshire State Model Educator Evaluation System". She asked people to look at the big picture while they read it. She then said that we would also look at very specific aspects of consequences and specific measurement approaches. She said if you agree with what you read, put a checkmark next to it. If you have questions you want to discuss, put a question mark next to that. And if there is something missing, you can put a "0". The room was divided into two groups. One group was to focus on consequences while the other group focused on specific measures. Kathy also told the consequences group to take a look at the notes from the last meeting (included in the folders) because there was a lot of conversation about consequences.

In relation to the latest iteration of the state model system, Scott said that if something is italicized, it is new. When it came to the consequences and support section, some of it was presented as a single document at the last meeting, but when looking at it, it did not make sense to put it all in one section. If there were pieces that some people wanted there and others did not, Scott put a strikethrough in the text. He added that if it was clear that people wanted certain things removed from the document, he took them out. Another key decision that would need to be made today is whether we use the word "shall" or "should" in the state model system.

Carol Kierstead

Carol divided the two groups (consequences and specific measures) into smaller groups for each topic. She then asked each group to go over the pages specific to their topic and identify three to four priority items (items of concern) with recommendations and put them on the flip charts.

After the group work, Carol brought everyone back together and the feedback from the groups is as follows:

SPECIFIC MEASURES

Group 1

1. Classroom Observations
    - What are we describing? — planned event or 'drop-in'; how extensive; four planned extensive observations are too many.
    - Two is enough for formal – with pre and post conference required.
    - Additional two (at least) for drop-in with formative feedback, in timely fashion!!! Feedback does not need to be lengthy, but needs to be timely and written with conversation as needed.
    - Additional peer observation and feedback should be formative and not evaluative.
2. Professional Portfolio
    - Need to have planning session (<u>with the time to do this</u>) to determine up front what will be in portfolio.
    - It needs to be individualized.

- The time should be spent to identify what artifacts will be used, not to build another portfolio – not extra work.
- It should be part of the existing beginning of year planning.
- Does it need to have artifacts for all five domains each year? We say, yes. But need to give the teachers the time. It should be integrated part of my work, not a big portfolio.

3. Professional Responsibility should be part of new teachers, too. It's a different type of PR, but still important. Being a good mentee, being an active participant in team meetings, parent relationships, etc.
4. Access to Test Data
   - Does not need to come from state if local district can do it on their own.
   - Page 6 Item 3, And/or. "*These results, based on NECAP and eventually Smarter Balanced Assessment Consortium (SBAC) tests, using the SGP model, shall be incorporated into teachers' evaluations either using a shared ~~or~~ **and/or** individual attribution framework."*
   - Also, on three. It should be "should be" not "shall".
   - Change 5, to say if multiple years are not available, then it can only be used in the context of other measure. *To the extent possible, yearly evaluations shall include multiple years of student performance results.*
5. Parent and Student Input
   - This should be included. Remove 'unintended negative…"
   - Must look for research based, well worded, tested tools.
   - Must ensure a high rate of return (at least x percent).
   - It needs to be a balanced component of the overall review.
   - Should not be a large percentage in itself.
   - Cannot be anonymous.
   - Survey needs to have supports in place for parents & students who have difficulty understanding or completing survey.
   - The use of the survey needs to be explained up-front to ensure accurate responses – to improve effective teaching and school climate.
6. Other Items
   - Why three categories not four levels of performance?
   - We don't understand the Panel examples.
   - Bottom page 4, add "formative". Including peer teams, in addition to building-level administrators, to participate in the formative evaluation process
   - Should there be guidance on % of teachers at each level. For example:
     - There should not be 90% of teachers highly effective.
     - It is understood that a rating of "effective" is a great thing.
     - You should not feel you are inadequate if you are not highly effective?
     - You do not need to have a bell curve.
     - Maybe "effective with distinction" rather than "highly effective", should be the top category.

Scott responded by saying that we need to be thoughtful about the surveys. If they count for any percentage, it can be tricky, especially when you get to student surveys. There could be serious implications with this.

Scott also spoke of peer evaluations and how they can be difficult. Having someone observe a foreign language class or an art class when they do not teach those subjects can be hard because they do not know the content and whether or not the teacher is teaching the correct content.

Group 2

3-4 Priority Items	**Recommendations**
• Student/Parent Surveys – Re: Evaluation (stated in document)	• ? of fairness and recency of events error
• Administrative Workload (the number of evaluations)	• District needs to equalize workloads. ? P.D. for evaluator
• SLO's (the group liked the idea of SLO's)	• Status • Conditional Status – Use of both • District-wide, etc. • Ind./Sub-groups
• Combining Multiple Measures (the group liked the idea)	• Establish the threshold(s) for each measure (fair and accurate)

Someone from the group said that the district they were in had done student surveys from grades 2-12. The lower grades had smiley faces to answer questions, rather than multiple choice questions. They discovered that timing was critical. When the surveys were given had a lot to do with what kind of information you got from it and the way the students reacted to it. They also discovered that in schools where the parents were allowed to pick their child's teacher (because the school was large enough), the results were very different from the schools where they had no choice in the teacher. The high school students were fairly objective to it and some good information was received, but again timing was important. We could get a lot of valuable information from parent and student surveys, but we need to be careful and do it over time, not just single snapshots. Give them a survey on the first day of school which asks them about the previous year, then you compare that to how they react about the ensuing year.

He also talked about timing and allowance for time. When we start looking at the mechanism, it is a great plan, but it would vary dramatically, in some cases, from year to year and school to school. For example, you could have a school that has all new staff members and the principal is overwhelmed, compared to a school that has no new staff members and a lot of experienced, well-performing teachers. We should establish some guidelines or give people some idea of how the proportion between the number of people that need to be evaluated at different levels and the number of support staff (evaluators) you need to get that done. He said that he has observed all types of classes and good teaching is good teaching, whether it is in French or chemistry. If the evaluator is properly trained and given enough tools, they can determine that. If there is an issue with content, that is another problem altogether. If you go into a French class, you can tell whether the students are engaged and whether the teacher is keeping them focused on the lesson. The key is to provide the evaluators with good models of training.

Scott responded by saying that people are actually pretty weak in observations with being able to discern instructional quality, but they are really good at being able to discern classroom management quality. There are people that are better at evaluating instruction, but it takes time.

Group 3

3-4 Priority Items	**Recommendations**
• Observation	• Needs to be defined

	▪ Formal ▪ Informal -Walk-through -Peer observation continuum -Videos -Peer sharing ▪ When is it/when is it not evaluative? • Define frequency • Differentiated, but not eliminated, observations for effective educators • Need to clarify the difference between evaluation and growth
• Simplify “Combining Multiple Measures” o Page 6: ▪ Compensatory ▪ Conjunctive ▪ Disjunctive ▪ Profile	• Reference page 8
• Consistency in Evaluation through P.D. • Clarify the terms Novice, etc.	• Delete 2 year window for Novice • Page 2, #13 contradicts SLO **(?)**

Scott responded by saying that he thought four observations for 45 minutes each throughout a 180 day school year (with the school days being 6-7 hours long) was a thin sample. On the other hand, given the number of teachers in a school and the amount of work that the principals have to do, it may be difficult to do more than that. This is something that will need to be worked through during implementation and different models may to be tried, rather than using the same old model.

Someone responded to Scott by saying that as an administrator who does observations, he knows what he wants it to look like. However, that is not necessarily the only way. We need to be clear on what the expected result is because no one has a full view of what good teaching is. There needs to be formative conversations with the teacher and then you can see where you get summatively at intervals (the end of that year, half way through that year, etc.).

Scott then asked for the state model if we should specify a recommended or specific number of evaluations or data collection protocols. Or should we leave it broader in that the districts shall determine a valid approach to collecting this data in these formats (which is vague)?
Someone else said it is not only about the frequency of the observations, but also broadening the scope of what that observation entails. She said her district is a SIG district and an observation is not necessarily a traditional conference to talk about what is going to happen, observe for 50 minutes, and then have a post-conference. It is a possibility, but it is also a possibility to do things differently. You can have evidence-based conversations to help that teacher grow.

The feedback presented by each of the groups working on Specific Measures prompted flip charts with “Implications for Decisions”. Following is the information written on the flip charts:

Implications for Decisions

- Consider changing "highly effective" to "effective with distinction"
- Student surveys ~ Elementary → High School
    - Could have unintended consequences

Consider how to frame this invitation to parents as a member of a learning organization.

A question of fairness.
*Timing is critical and allowance of time.

Strategy…Give survey at beginning of year and ask students to "rate" former teacher, then after a few months, rate current teacher and compare how students respond.

*Choice in teacher selection matters.

- What about peer review?
- # of Evaluations – Needs to be realistic and there should be PD for evaluators

Need to prepare/train evaluators how to be effective observers of instruction as well as classroom management.

- Definition of "observation" is critical, e.g. formal, informal, walk-through, etc.
    - Implications for what observation "looks like", e.g. evidence-based conversations
    - # of observations critical – could be differentiated, e.g. novice vs. experienced

---

CONSEQUENCES

Group 1

3-4 Priority Items

- Differentiation

Recommendations

- Eliminate #1
- In #2, change wording to: "New educators shall be evaluated yearly until they reach professional status."
- In #3, change the wording to: "Teachers with professional status shall be evaluated every 2 years until they receive highly effective ratings."
- In #4, change the wording to: "Teachers with highly effective ratings shall be evaluated every three years."
- In #5, change "partially effective" to

	"needs improvement". • In #6, change "evaluations" to "mentoring".
• Consequences	• In #1, change wording to: "An experienced, professional educator with 2 consecutive years of ineffective ratings may be non-renewed (pending the unsuccessful completion of a professional improvement plan)." • Eliminate #2 – Mentioned in Differentiation • Eliminate #3 • In #4, change "should" to "may" in line 1 • Eliminate #5 • In #6, change wording to: "Educators rated in need of improvement based on 5 Pt. system may be non-renewed."

• Categories	• Move from 4 Categories:
	  ▪ Highly Effective
	  ▪ Effective
	  ▪ Needs Improvement/ Developing
	  ▪ Ineffective
	• To 5 Pt. Categories:
	  ▪ Highly Effective
	  ▪ Effective
	  ▪ Developing
	  ▪ Needs Improvement
	  ▪ Ineffective

This group stated that they liked the identifiers of highly effective, effective, needs improvement and ineffective. They did, however, have a hard time with needs improvement because there are going to be teachers that do not fit into that category because they are moving along and headed for effectiveness, but they are trying something new, shifting grades, etc. They are wondering if a fifth category should be added to identify those teachers. Scott said that was something that would be talked about later in the day. They asked Scott if there was any reason that a 5 point system would not work, rather than a 4 point system. Scott said the problem is with having enough information to cover the 5 points. He also said that he had a hard time with "developing" and "needs improvement", as did that group. He said that they are at about the same level, but one implies that one is on their way out and the other implies the same thing, that one is on their way out. Someone else said that they are very different based on the circumstances. Scott said that one of the ways to handle this would be that the first year that you are in that category, you are "developing", whether you got there by going down or by going up. And then if you are in for a second year, you are "needs improvement". However, this could add a level of confusion to the communication plan. Someone else suggested that "developing" be used for people that are new to the profession or are switching assignments and “needs improvement” be used for people that have been in the profession for a period of time.

The group had no issues with the “Assumptions” and “Support Systems” sections of the document. When they looked at the “Differentiation” and “Consequences” sections, they felt the priority should be on simplifying it and making it coherent with our legal requirements.

<u>Group 2</u>

<u>3-4 Priority Items</u>	**<u>Recommendations</u>**
• Supports o Concerns about where the money is coming from for the training o Concerns with the panel chart ▪ Are the “pro practice” and “student perf.” numbers district-developed? o Concern with the language ▪ “Licensed personnel” vs. “educator”? ▪ “Ineffective” vs. “partially effective” ▪ “Should” and “shall” o Concern with #1 ▪ Seems like #1 is trying to address 2 separate issues: Awareness and Process & Training for Observations o Concern with #2 ▪ What is the “well-specified system” vs. induction and mentoring? ▪ Professional communities o Panels → Weighting student performance and professional practice → a district decision?	• Either articulate more or take away • Replace “should” & “shall” with an action verb. Ex: “A district develops” instead of “A district shall develop" • Consider separating out the issues
• Differentiation	• #1 - “does not preclude formative (continuous) feedback…” • #2 - Legal – 5 year vs. 3 year
• Consequences	• #1 – “current status” → More specific language: ▪ i.e. Supports, improvement plan, awareness plan

	▪ Supportive language but no surprises • Differentiated system of supports evaluation based on experience ▪ Better than the 2 year, 3 year, 5 year language • "Reward" system → "Should decide in consultation with administrator on mutually agreed upon incentives that better the school community."

Group 2 stated that they had focused in on legal issues, language issues, examples and specific vs. general. Someone commented on the panel chart stating that it looks like you have three different systems, three different sets of language. The vertical axis is Danielson, which is "distinctive", "proficient", "basic" and "unsatisfactory". It is confusing how those four categories with those four separate ratings feed into a single rating and whether or not we want to specify that at this time. Then on the bottom, "Student Performance Rating", we have "1", "2" and "3", they came up with "not met", "met" and "exceeded". How does that come about because those are the details that people want to know? How do those two axis feed into that one? This whole system leads into a fourth system that deals with continuing contract and status as an educator and that needs separate language. At the beginning of the document it says that these are going to be four equally weighted areas plus one for about 20%. Then later in the document there is a whole different weighting system.

Scott responded by saying that it is just an example. He said if you go through Danielson, there are dozens of rubrics. The question is what to call each thing. Ultimately we need to come up with a single rating for each teacher and we need to decide if it will be a local decision or not. If you think about Danielson, do you need to be proficient in every rubric rating to be a proficient educator? Scott said he would hope not. He said that the document contains one example, but there are other ways to do it. The 1-4 ratings on the vertical axis are comprehensive evaluations of the four domains. We still need to figure out how you get there though.

Group 3

3-4 Priority Items	**Recommendations**
• Assumptions	• In #1 change "figure out" to "guide" • Add a #3 – A statement that the process is documented
• Supports	• Throughout the document, it says "research-based" and the group recommends changing it to "research and evidence-based". • In #1- Break the 1st sentence into two. ▪ Have the 2nd sentence include support, consequences and rewards. • In #2 - Mentoring must be a formalized process

	• In #3 – Ineffective and on improvement plan – Ineffective in one area/domain? Or overall? Get rid of “partially”. • #4 & #5 – Strike
• Differentiation	• #1 – Eliminate • #2 – Add “2 yearly evaluations”
• Consequences	• #2 – Eliminate • #3 – Change it to say “…three consecutive years in the same position of developing…” • #5 – Eliminate • #6 – Eliminate

Someone had suggested that wording be added so that teachers do not habitually switch positions in an effort to protect themselves from having to improve. We do not want to create a system that enables people to continually switch their positions.

Someone else said that we need to be very clear on the terms “dismissal” and “non-renewal”. It is a very important distinction because “dismissal” sounds like you are leaving immediately, whereas “non-renewal” sounds like you can finish the school year but you will not be returning the next year.

The feedback presented by each of the groups working on Consequences prompted flip charts with “Implications for Decisions”. Following is the information written on the flip charts:

Implications for Decisions

- Performance levels – Need to consider how to balance Needs Improvement and Developing ~ Do we need a fifth category?
- Consider using term “New” vs. “Novice”. Will continue annual evaluations until they reach professional status and professional status teachers will be evaluated every 2 years until they reach Effective rating, then every 2 years after that.
- Only educators with Effective ratings could participate as mentors, PDers.
- Supports ~ Where will funding come from?
- Need more clarity Re: Panel Chart – e.g. Page #9 (#1 on Supports – Seems like 2 issues are being tackled – separate out); in-effective vs. partially developed; take out “should” and “shall”; need to balance (and be clear) supportive and accountable language.
    - Create flow chart Re: # of years – ineffective to effective – or not
- Issue of rewards could be defined broadly beyond only $.
- #1 Assumption - Δ to guide
- #2 Assumption - add how process is documented
- Supports – Δ from “research-based” to “research & evidence-based”
- Supports – Sentence #1 – Break into 2 sections
- Supports – Sentence #2 – Say mentoring is formalized
- Supports – Sentence #3 – Need to clarify what ineffective really means

Carol Kierstead

After the group work was discussed Carol facilitated a discussion to begin making some decisions. Carol said that one of the items that has been discussed a lot is the performance levels. She said that we need to determine what the purposes are for the creation of the performance levels. What do they have to accomplish? What are the criteria that they have to meet for us to be able to say that we created a set of performance levels that are going to do the job and what is the job that they have to do? After everyone thought about this individually, they discussed some of the criteria that they thought about. Each group then shared one of the criteria that they came up with. Below is the information recorded on the flip charts regarding the criteria.

Criteria for Performance Levels

- Provides common language with understanding of the language
- Differentiate performance
- Identify teachers for support, non-renewal, or leadership
- Encourage continuous learning, reflection, communication and sharing
- Label for high-stakes performance decisions and formative, collaborative improvement
- Clearly communicate to teachers and leaders expectations for growth as a profession
- Recognizes achievement
- To inform professional development
- Help develop clear pathways of teacher improvement to improve student performance
- Have effective teachers in all New Hampshire schools
- Provide summary descriptor based on multiple sources of data
- Reference to descriptors in PLD (Performance Level Descriptors)
- Identify if teachers are on the right track

Mike Schwartz

Mike went over a PowerPoint presentation which explained the $5,000,000 Longitudinal Grant that we just received from the USED. The intention of the grant is to develop data systems for states. While the grant can be used for early childhood, K-12 or post-secondary education, New Hampshire chose to focus the grant on K-12. There are three primary outcomes of the grant that we are trying to achieve. The first is focused on the evaluation of effective educators. The second is focused on informed decision making and the third is focused on College and Career Readiness.

Mike then went over the outcome in more detail. In the grant they had to define, very specifically, the different outcomes that they wanted to achieve and the deliverables associated with those outcomes. He did state, however, that there is flexibility to adjust these outcomes to better fit New Hampshire. They came up with four main deliverables in the evaluation of effective educators:

- Develop inventory of evaluation system data elements and integrate into SLDS (e.g. teacher roster, assessment data, competency data, survey data, etc.)
- Streamline District Data Submission (reduce effort for district)
- Develop web tools to allow supervisors to view evaluation data
- Create a system to administer the evaluation process

Mike said that we are in the process of including a growth percentile model in PerformancePLUS that the Center for Assessment is assisting with. It will allow the school district to print a roster for each teacher and their students and the growth percentile for their students. The intention is to help the teachers better understand who the students are in their class, who is having trouble in certain areas, etc. There are four potential growth models that we can use or build off of:

- P+ and EIS
- LEA Model that leverages P+ for student outcomes

- Other state-wide growth tools
- Fully LEA based model

Another big piece is professional development and support and there is money in the grant for that. One of the components of the grant was to develop this idea of an online social learning network to connect schools for collaboration and training. There could even be learning communities around certain topics, such as student growth.

The EIS system could be expanded to track the evaluation process of the teachers in the district. The PerformancePLUS system could be expanded to allow viewing of student growth on NECAP, as well as other national and local assessments, parent surveys, etc.

Someone asked Mike how we ensure that this information is kept private or only goes into the correct hands (using the example of the teachers in New York that are labeled the worst teachers in the state). Mike said that right now the system has security role rights that are accessed at the local districts. We can certainly create reports that the principals and superintendents can run to see who has accessed the system. Mike said that we need to look further into this to protect the information. Judy Fillion then confirmed that the rules govern what is allowed to be released regarding the educators in New Hampshire.

Mike ended his presentation by handing out a questionnaire for people to answer in order to gain input from the field. It contained the following questions:

- What are you using to track evaluation status?
- What are you using to track completion of evaluation components (e.g. completion of classroom observation, growth reports, etc.)?
- What are you using to identify student growth results?
- Do you need tools to manage classroom observations, and other Charlotte Danielson *domains*?
- Where do you see the biggest need for data system support?
- Would you be interested in being part of an advisory group to help define these tools and support?
- Other comments, suggestions.

Carol Kierstead
Carol said that the group was now going to come to decisions regarding the first two items on the decision sheet in their folders. We need to decide how many levels we are going to have and what we are going to call them.

Someone said that you are either effective or you are not and then asked how long children should have ineffective teachers. They further suggested that there are basically two tracks for the educators to be on. The first track is for those educators that have been basically doing the same sort of assignment for many years. You then have those educators that are just coming into the profession or who, for whatever reason, have had to change what they are doing and learn new skills. You cannot expect them to just jump into something new and be effective and this situation requires a non-threatening, very supportive category that acknowledges that. On the other hand, the educators that have been doing the same assignment for many years should not be ineffective for more than a year. There can be difficult periods for all educators, but after a year they should be back to being an effective teacher. Carol said that there are two issues with this. The first is what you call an educator at a certain point and then what you do about them based on what we call them. She then said that we first need to figure out what to call them based on who they are and their status in an educational system. It was then suggested that we call them an experienced educator. Carol and Scott agreed that calling someone experienced is less about performance and more about how long they have been in the profession. Scott then asked if we should have different classifications for "new" teachers (and we would, of course, need to define what "new" means). Someone agreed that we would either have two different groups of performance levels or have the five categories that have been talked about. Developing is different from partially effective or needs improvement. Carol then asked if you get to having to describe the rubric for a set of performance standards and indicators, (what those two distinct categories mean) would you be able to create that? Someone said that they would not be based on the criteria for performance levels that were identified earlier that day. If the purpose of the performance level descriptors is to differentiate, it does not matter whether you have been on the job for two years or twenty years. Carol then asked if the struggle was with what we were calling it, and not with how the level of performance is described.

Someone said that a new teacher would go into the developing category if you are using highly effective, effective, developing, and ineffective. He then said that no one can go from effective to developing. If you are an experienced educator and you are effective, you cannot go to developing. The only people that can go in the developing category are new teachers. It was then agreed that an exception to that would be someone switching assignments, but that you could not take a step down due to performance.

Another person stated that a teacher is either effective or they need improvement and it does not matter if it is their first year or not. Someone else asked that if a brand new teacher walks in the door and they are the best teacher around, will they still be placed in needs improvement? It was answered that anyone can need improvement. Someone else made the argument that everyone has been in the developing category recently because of technology and it has nothing to do with effectiveness in many other areas of teaching. She said that all educators should be developing in some way at any given time.

Someone then asked to have the difference between needs improvement and ineffective clarified. Someone said if you are ineffective, you are not helping students move forward. If you need improvement, there is some hope that you can get better.

Some people said that it does not matter what you call them. Other people said that it does matter what you call them because the language has power and it conveys an attitude. There are a lot of educators that are in the process of developing and moving toward effectiveness and want to feel supported. To say that they need improvement is wrong. We need to use language to support the teachers, rather than language that would take away their confidence.

Someone then asked what you call an effective teacher that, no matter what the circumstances are, they fall down. That was answered with ineffective.

The real issue is whether we want to have a four point system or a five point system. If we do the five point system, we describe that fourth category as developing or needs improvement.

It was then brought up that more important than the names of the performance levels is the descriptions of what those levels will look like.

There is also a concern about what happens when there is a change in administrators. One administrator may view an educator as effective, while another administrator may view the same educator as ineffective. For that reason, needs improvement should remain as a level. Someone else then stated that if there is a new administrator that does not believe the teacher is effective (as the previous administrator did), then maybe you get placed in the category of developing for the first year and if you do not improve after that first year, you get placed into the category of needs improvement. Someone then suggested that if a teacher is on their way up, they would be developing and if they are on their way down, they would be placed into needs improvement. It still provides an appropriate middle ground. Kathy said that the conversations regarding the evaluations will occur differently based on whether the teacher is experienced or new and what performance level they are on.

Kathy then brought up an item that the group appeared to agree upon. She said that everyone agreed that there are different reasons why people need to get better, that are not necessarily in the effective category (whether they are new to the profession, new to the content, etc.).

Scott Marion
Scott asked the group if there would be common descriptors like the examples in the document for the four levels. Will those be part of the state model system and, if so, what is the expectation of districts? Someone asked how many descriptors there would be and Scott said there would be four, to describe the names of the performance levels that are decided on. Scott said he believes that the state model system needs to have them and then we need to figure out what is required of the districts. It was then clarified that the state model would have paragraph descriptors and based on those descriptors you would then build a rubric or decision matrix to that.

Someone else suggested going back to the work that has already been done. We have already defined what effective teaching is. We could recommend that districts use existing rubrics that are already out there to save time and work. Scott agreed that effective teaching has been defined, but highly effective, ineffective, etc. have not been. Scott said that it may benefit everyone to have some commonality, rather than have schools and districts using different rubrics.

Scott then asked the group if they agreed that the state model should have at least four levels of performance. The group agreed with that. The flip chart indicated the following number of votes from the group in regards to how many levels the state model should have:

- **State model will have at least 4 levels of performance:**
  - 4 levels = 19
  - 5 levels = 5
  - 4/5 levels = 2

From that came some possible names for the performance levels. The following names were written on the flip charts as suggestions:

- Highly Effective/Effective with Distinction
- Effective
- Developing/Needs Development/Needs Improvement
- Developing
- Needs Improvement
- Ineffective

It was also brought up that the emphasis seems to be around teachers being effective with the challenged students. Effective teaching is not just engaging and educating those students that struggle, it should be across all levels. If you have gifted students, you need to be able to keep them engaged, as well, or you will lose them. Effective teaching needs to cover that level, too.

Someone asked how hard it would be to do all the categories. An example of a PLC was given and talked specifically about an example of an effective teacher. From that we should be able to determine what would describe the other performance level names. We could have a complete matrix of sixteen categories that could be used as models for districts to adopt. How hard would it be to differentiate in those areas by 1, 2, 3, and 4? Kathy said that if you look at Danielson, there are 22 components and multiple elements. Using Danielson as an example, you have the domain level, the component level and the element level. The purpose of that is for the formative and then, ultimately the summative.

There was some discussion about what to name the fourth performance level. Would we go with highly effective or effective with distinction? Someone said it depends on what your expected outcome is. If you want it to be an attainable, realistic goal then you should go with highly effective. If you want to separate out the top 5% of teachers, then you should go with effective with distinction. Do we want people to strive for it and see it as a reasonable goal? Kathy said that we will continue to have the dilemma of naming it until we get it defined.

Scott said that we did get agreement on some of the levels, but we will not get them all until we get the descriptors. Scott said that working on those pieces can be done virtually for those that want to assist.

Kathy Dunne

Kathy then let everyone know that they would be split into groups of no more than four or five. She then asked the groups to look at the decision sheet in their folders. The focus was going to be on questions 3 and 4.

For question #3 the groups needed to decide if the frequency of required evaluations should be identified as a part of the state model system. If they answered yes, then they needed to determine the frequency of the required evaluations for the following categories of teachers:

a. Effective/Highly Effective continuing contract teachers?
b. Needs Improvement continuing contract teachers?
c. Ineffective continuing contract teachers?
d. Novice teachers?

For question #4 the groups needed to decide if the consequences of ineffective ratings should be identified as a part of the state model system. If they answered yes, then they needed to determine the consequences and whether or not they agree with them for the following categories:

a. 1 year = additional support?
b. 2 years = dismissal or more support?
c. 3 years = dismissal?

After the groups had their discussions, the decisions were shared and recorded on flip charts. Below is what was recorded for each group:

DECISIONS

Frequency of Required Evaluations (#3)	**Consequences of Ineffective Ratings (#4)**
• Include? – **Yes** a. Minimum of 1 every 3 years - more if needed b. One yearly with 2-3 observations c. One yearly with 2-3 observations d. One yearly with 2-3 observations	• Include? **– Yes** a. Agreed with this b. 2 years = directed support by supervisor c. 3 years = dismissal
• Include? – **Yes** a. One every 2 years for effective – One every 3 years for highly effective b. Yearly w/ intensive interim observations	• Include? – **Yes** a. 1 year = additional support with targeted PD plan b. 2 years = non-renewal

c. Yearly w/ intensive interim observations d. Yearly w/ intensive interim observations	
• Include? – **Yes** a. One every 2 years for effective – One every 3 years for highly effective b. Yearly w/ intensive interim observations c. Yearly w/ intensive interim observations d. Yearly w/ intensive interim observations	• Include? – **Yes*** a. 1 year = mentor support and improvement plans b. 2 years = If they are really working on the improvement plans, they will receive additional support for one more year. 2 years = If they are not working and not making great improvements, then they will be dismissed. c. 3 years = If the teacher is given the third year to continue to improve and they do not, then they will be dismissed. *The group did not really get to talk too much about #4, but Cindy Chagnon said the above is basically what her district is doing.
• Include? – **Yes** a. One every 3 years for highly effective b. One every year for needs improvement c. One every year for ineffective d. One in the first year for novice and then dictated by your status	• Include? – **Yes** a. 1 year = additional support b. 2 years = more support c. 3 years = there is no third year if at the end of your second year you are still ineffective
• Include? – **Yes** a. One every 3 years for highly effective and effective b. One every year for needs improvement c. One every year for ineffective d. One every year for novice	• Include? – **Yes** a. 1 year = additional support b. 2 years = non-renewal

A concern was raised about questions 3 and 4. Someone had said that the consequences should also be put in place for other non-favorable ratings. Scott said that what could be done is to create different levels or types of support based on the rating that you have (novice, ineffective, needs improvement).

Scott Marion

If you are in grades 4-8, you have NECAP test data. If you are in the second year of that (the teacher of record), you can calculate student growth from the data. Scott stated that we could put in the state model system that the test data needs to be a part of the evaluations (it could be a percentage). Or we could put in the state model system that we are leaving it to the local districts and say that it is required to be used in at least shared attributions. The shared attribution could be based on the grade level or the entire school.

Scott then posed a question to the group. In the state model plan, if there is state test data available (and there always will be), does the state test data count in the teacher's evaluation, either as an individual or a shared attribution? This is question #6 on the decision sheet and the groups had conversations surrounding this question. The flip chart indicated the following number of votes from the group in regards to question #6.

<u>Use of State Test Data? (#6)</u>

6a. Required as part of the individual evaluation for all teachers in "tested" grades and subjects?

- Yes = 11
- No = 13* (*One group thought that individual teachers should have the option of using the test data.)

6b. Required to be used in the evaluation of teachers in at least a shared attribution framework?

- Yes = 23
- No = 1

Someone then asked what the percentage of our teachers fit into the "tested" grades and subjects. Scott said he was not exactly sure, but in most states it is between 25%-30% of teachers that fit into that category.

Scott then said that it is now time to focus on the other 70% (+ or -) of the teachers. This is related to the use of SLO's (question #7). He posed the following question to the group. Are student learning objectives (SLO's) required of everyone or just teachers in "non-tested" subjects and grades?

One school is using SLO's for everyone and doing some shared attribution of state data, as well. The SLO's that are being written have language taken directly out of the Common Core, and they are looking toward Smarter Balance. There is quite a bit of cross-over of the already existing state and local measures that they are using. They are keeping them really skill based and they have to write 2-3 of them. Two are part of their goals that are linked to their professional development and recertification, so it is all going to be one big system. Their whole teaching goals are tied to the SLO's. Each person needs to have 2-3. Some of them may be grouped. But if you are the stand-alone French teacher, no one else will have those specifically, but you can certainly share it across the curriculum. Scott then said that as we move towards competencies, they are the ideal target.

The groups then had a discussion about question #7a. The flip chart indicated the following number of votes from the group in regards to question #7.

<u>Use of SLO's (#7)</u>

7a. Required as part of the individual evaluation for all teachers in "tested" grades and subjects?

- Yes = 24
- No = 0

Scott said the group was not going to talk about the last question (#8) because there is not really a decision to be made. Question 8, letter c is off the table and the distinction between a and b is not clear enough right now.

Scott said that he would like to get the group back together one last time before September. Anyone who would like to be involved in the group that is creating the descriptors can e-mail Scott. In addition, anyone who would be interested in being a "reader" to go over the document can let Scott know.

SCEE Team
Ashley Frame gave a quick overview of the SCEE team conference in Seattle. It provided an opportunity to hear what other states have done, what challenges they have faced, what has worked, etc. She said Delaware had some issues with their SLO's because they did not make them measurable or smart and they wrote them for their teachers. It was good to see some mistakes from other states so we can be more aware of what not to do.

Randy Bell then said that they saw a presentation from Ramona Pierson, a neurologist and entrepreneur. She has done a lot of research and she spoke about real 21st century things (maybe even 22nd century). Randy also said that last year's teacher of the year spoke about the complexity of teaching, which solidified for Randy some of the things that we should be doing and some of the things that we should not be doing in terms of evaluations. He is concerned about them being content oriented. He said that Linda Darling-Hammond's presentation was excellent as well.

Commissioner Barry
Commissioner Barry thanked everyone for all of the work that they have done throughout this entire process. She also said that certificates will be given out for the hours that the task force has put into this. She also announced that Rhonda Wesolowski would be retiring and thanked her for her help.

The meeting ended at 4:01 pm.

Submitted by Trisha Allen

July 26, 2012