



Joint Meeting
Commissioner's Task Force to Develop a Performance-Based
School Accountability System and the
NH DOE Accountability Task Force
October 14, 2010, 8:30 am – 12:00 pm, Room 15

NEXT JOINT MEETING:

Monday, November 1, 9 am – 12 pm, Board Room

Attending:

X	Virginia Barry, Ph.D.	Commissioner of Education
X	Brian Cochrane	Director of Assessment and Accountability Nashua School District
X	Paul Couture	Principal, Stevens High School, Claremont
X	Jerome Frew	Superintendent, Kearsarge Regional School District
	Molly Kelly	Chair, Education Committee, NH Senate
X	Daphne Kenyon	NH State Board of Education
X	Paul Leather	Director, Division of Adult Learning, NH Department of Education
X	Scott Marion	National Center for the Improvement of Educational Assessment, Dover
X	Judith Fillion	Director, Division of Program Support, NH Department of Education
X	Edward Murdough	Bureau of School Approval
	Kathleen Murphy	Director, Division of Instruction, NH Department of Education
X	Emma Rous	Chair, Education Committee, NH House of Representatives
X	Vincent Spiotti	Bethlehem School Board, Bethlehem, NH
	Franklin Gould	NH House of Representatives, Lebanon, NH
X	Deborah Wiswell	Bureau of Accountability, Curriculum and School Improvement, Division of Instruction, NH Department of Education

Guests:

Dean Michener, Executive Director, NH School Boards Association

Accountability Task Force (in addition to those listed in the Commissioner's Task Force)

District Representatives: Helen Bickford (SAU 53) Patrick Connors (SAU 53Epsom); Heather Cummings (Gov. Wentworth); Donna Crook (MSD); Kathy Stavenger (SNHU), Steve Zadavec (Portsmouth)

NH DOE: Sally Fellows, Steve Bos, Tim Kurtz, Susan Randall, Gaye Fedorchak, Ginny Clifford, Merry Fortier, Ed Hendry, Cathy Higgins, Keith Burke, Marcia McCaffery

Center for Assessment: Scott Marion
Measured Progress: Shannon Douglas

NOTES:

Meeting Objectives:

- ◆ Examine statewide and selected sample school data for SGP using Met/Not Met chart
- ◆ Review sample score reports using proposed weighting
- ◆ Reach final decisions on indicators, weighting, and format for reports

1. Deb opened the meeting at 8:40 am. Attendees introduced themselves, their task force role, and their desired outcomes for this meeting.
 Tim Kurtz expressed his appreciation for the fact that the accountability system being developed by these task forces is a *state* designed system. It allows New Hampshire to define what 'we' want to hold ourselves accountable for. He expressed the hope that when this system is in operation, there will no longer be reference to what "they" expect us to do, but rather to what "we" think we are responsible for.
2. Deb distributed and reviewed the *NH Accountability System Overview* document (attached) which describes the purpose, components and outcomes of the NH Accountability System being developed. Deb described the progress so far with the Input component. Over 25% of schools have submitted. Principals Paul Couture and Pat Connor reported that the submission process was labor intensive, but that the system was easy to navigate. The major decision yet to be made by the Commissioner is what will be acceptable as "adequate."
 The performance based system being created by the Commissioner's Task Force includes two parts – *state defined indicators* and optional *locally defined indicators*. Recently, the Commissioner requested an extension from the legislature for the required November progress report on the design and testing of the performance-based system.
3. Deb distributed the *Key Terms and Examples* handout with definitions of the terms used in discussing the growth model based on student growth percentiles. The handout includes a set of examples for students and schools, describing what the percentiles mean.
Discussion:
 - Tim clarified the meaning of "median" in the glossary.
 - Heather suggested that language should be adjusted on handouts for parents and general readers.
 - Emma asked and Scott confirmed that the SGP is calculated based on all students in the state who took the same assessment (not just the students in their school)
 - Steve Bos recommended using graphics to improve the explanatory value of the handouts.

-- Gaye recommended adding "across the state" to the *academic peers* definition to highlight that students are compared to all others who took the same assessments.

4. Scott presented the statewide data to inform decisions by the Task Force on the elementary and middle school performance based measures.

(Slide 2) Scott reviewed the decisions to date on: subgroups, minimum N, ELL performance, participation rates, "small schools" (i.e., K, K-1, K-2), high school indicators, and content areas included (reading, math, science, writing),

Deb distributed a handout describing how ELL performance is measured in the state. Gaye Fedorchak clarified questions. Paul Leather pointed out that a state task force is in the process of developing an accountability system for ELLs. Scott and Deb proposed that the Department not define multiple targets for ELL. Paul asked that the performance based accountability system not be finalized until the ELL Task Force has an opportunity to consider all options. Scott reminded everyone that the Performance Based Accountability System has a short timeline to begin piloting. Scott and Paul will continue discussion of the ELL accountability issues outside this meeting.

Participation rate (slide 3): Scott reminded members of the existing federal requirements for 95% participation in state assessments. NH data shows most schools already meet the 95% target. Task Force members recommended including participation as a factor in judging adequacy in part to show where there may be gaps in participation for the subgroups. For participation, N = 40 as in the federal system, aggregating across years if needed to reach the minimum N. This contrasts with the minimum N for other indicators proposed for the Accountability System.

Bubble Charts (online): Scott showed the bubble charts for mathematics with subgroups in different colors, and bubble size based on the size of the subgroup. Some observations –

- SWD (IEP) students demonstrate generally lower achievement
- Whole school bubbles are more widely distributed, many more clustered at the upper half of the vertical axis (percent proficient).

Bubble charts are descriptive and helpful but not ideal for aggregating across all subjects and subgroups. Previous discussions have examined the options and challenge of setting standards for "adequate" or "not adequate" by using the bubble charts alone. (See previous meeting minutes.)

One option that could work is to use the student growth targets which are calculated for each student. Growth targets address the concern that assessing absolute growth is insufficient since students may grow at a high rate but still be below proficient. After extended discussion over previous meetings of both Task Forces, a decision has emerged to establish a target of meeting or maintaining proficiency within three years or by 8th grade, whichever comes first.

School growth targets (slide 6): The Performance-Based Accountability System could use student growth targets as an indicator of adequacy. For example, the system could aggregate targets for every student in a school, calculate the median target, then compare the median target with the median actual/ observed SGP for all students. This comparison would answer the question: did this *school* meet or not meet its target? One option is to define 'adequacy' as reaching a meaningful target.

M-T bar charts (slide 8) show the information for individual schools where M is the median observed SGP, the T is the median target SGP, and the length of the green bar shows the percentage of students in the school who met their target. Scott explained that schools with

relatively lower performing students will have high targets, so it isn't defensible to use this chart alone as the measure of adequacy. Looking statewide, many "whole schools" meet their target. The picture changes for subgroups.

(slides 9 & 10) Scott reviewed the "rubric" approach for assigning points toward adequacy for the five subgroups and four subject areas proposed to be included in the performance--based system. The rubric allows points to be awarded for each of the four subject areas for each of the five subgroups. Deb explained how the rubric system would be applied to NH schools. She described what the actual state data looks like. The distribution of rubric points across the state seems reasonable and defensible. Schools earning the lowest points are those who demonstrate significant challenges in achievement on other existing measures. Those earning high points on the rubric demonstrate high achievement in other measures.

(slide 12) Scott showed the percentages of schools in each subgroup for each rubric score. This data can answer whether the use of this rubric "makes sense" and "seems fair." Very few schools who did not meet their target earned 4 points (highest). Very few schools that DID MEET their target earned only 1 point.

Deb distributed the sample elementary and middle school results for SGP for each subgroup (except ELL – data not available yet). Points are allocated using the rubric for each subgroup. Comparing actual schools with similar percent proficient scores on the bubble charts and rubric values, demonstrates the system credits schools for high growth while recognizing less strong percent proficient.

Questions and discussion –

Q: Can it be described as the "minimum" target, to be better able to explain how they could meet their target but earn only few points.

Q: Is there any value to create curved "bands" on the bubble chart and assigning points in that way?

Scott responded that the band system was proposed in previous meetings and considered. The challenge is to define a defensible rationale for the placement of the lines defining the bands. The proposed rubric system gives credit for moving students forward at a high rate AND having high rates of proficiency at the same time.

5. Scott opened a discussion of science scores in the elementary and middle school (slides 15, 16). No growth is available for these two subjects, only "% proficient" (slide 15). Scott showed a chart with state data for science, with the mean, median, and scores at the 25th, 50th, and 75th percentile.

Based on the statewide data, Scott proposed awarding points by (a) 0-20% = 1 point; (b) 21-40% = 2 pts; (c) 41-65% = 3 pts; and (d) 66-100% 4 pts. The remaining challenges are whether to assign high points to average proficient percents of 50, which is not generally commendable.

Discussion:

- propose 0 to 40 as earning only 1 point; rationale is that, if the test is fair, then students should be achieving proficiency; would result in large numbers of schools earning fewer points
- why not do an index score ? Index values would give the school credit for "high" 2s (students almost proficient)
- if we change this indicator to an index value, will need to consider consistency because SGP uses percent proficient rather than index on the vertical axis of the bubble chart; the differences in results would be minimal

- when students need to make up ground academically in reading and math, often must take time from science so a lower percent proficient might be understandable (but should we support that?)
- bottom line question for assigning values is what would be evidence of NOT providing an opportunity for an adequate education?
- caution not to have the system encourage schools to focus on only a select group of students;

Scott referred to the actual science values when state data is examined (slice 17). There are a high percentage of schools that earn only 1 point for their ELL and SWD groups, and 40% of schools earn only 1 point for their low SES group.

Decision Question: Do we count ELL for science? To be decided.

6. Writing: Scott referred to writing scores statewide. Like science, no growth can be calculated for writing, only % proficient and above. (slide 18) The table shows mean, median, and range for each subgroup. Scott proposed (a) 0-30% earns 1 point; (b) 31-50% earns 2 points; (c) 51-65% earns 3 points and (d) 66- 100% proficient earns 4 points. The distribution of schools across these four point groupings (slide 20) are similar to the science distribution.

Questions to consider: are these cut points 'right'? Do they make sense? Do we count ELLs for writing?

7. Scott reviewed the array of indicators and points to be earned toward the final score towards an adequacy decision (slide 22). Still under discussion is the "attendance" rate using a measure of students who miss more than 10% of enrolled time. Mike Schwartz is running state data to see the distribution of students missing 10%, 15%, 20% and 25% of eligible enrolled days. Data will be reviewed at the Nov. 1 meeting of the Commissioner's Task Force.

Discussion –

- don't want to "double" dip and penalize students who are ill or absent for reasons beyond the school's control
- caution on creating a new definition of attendance that doesn't currently exist; a state definition for truancy exists, but refers to "unexcused" absences, which each district defines for themselves; caution on complicating the conversation by adding new measures
- attendance is currently part of the federal accountability system
- suggestion: look at the actual data and decide if the measure creates the incentives desired for the system
- Donna reported that a new system implemented this past summer in Manchester has resulted in higher attendance rates
- concerns about alignment between truancy laws and this system
- Ed Murdough clarified new state definitions regarding the number of instructional hours, number of days, etc.

8. Scott initiated a discussion of weighting of the various ES-MS indicators (slide 23). Deb distributed a sample report summary form to show the data that would be available for each school. By looking at the reports distributed today and filling in the points earned for each subject, subgroup for each school, a tentative ratio score emerges. Reports will also include a bubble chart and M-T line graph which are interactive for each of the four subject areas and subgroups. This supports the recommendation made earlier in the meeting to include visual graphic representations of the data.

Discussion

- Keith's bubble graphs are in development; he will try to have filter-able fields such as "met/ not met" to all the bubble graphs to be filtered by various fields

- Deb announced that, since a grant funding further development of the system was not received, the Department will develop the bubble charts and reports on their own (working with Damian from the Center and Keith Burke); in the early stages, the system may not have all the ideal features desired, but they will offer a useful starting point
 - Gaye offered a suggestion that reading and math count 4X science and writing; this takes into account the greater number of tests given in reading and math at the elementary and middle school grades
 - Brian suggested 3X weight for reading and math over science and writing; also recommends weighting growth more than status;
9. Deb reported that the Commissioner requested that these meetings continue as Joint Meetings of the CTF and AYP/ATF groups at least until the performance-based system is fully designed.

TO DO:

1. Complete calculation of attendance/ truancy data.
2. Prepare actual score summaries for selected high schools.
3. Run scores for all schools in the state using weighting of 3X and 4X for Reading and Math over science and writing and other weighting
4. Report from the state ELL work group on their discussions.

Preliminary Agenda for next meeting, Monday, Nov. 1 --

1. Attendance how to report, include or not
2. High School indicators
3. Reports for all ES- MS with different weighting options
4. Review participation values

UPCOMING MEETINGS (ALL WILL BE JOINT CTF/ATF)

Monday, November 29, 9 am – 12 pm

Thursday, Dec. 16, 9 am – 12 pm