

AQuESTT:

Accountability for a Quality Education System, Today and Tomorrow



AQuESTT Performance Classification Task Force Report

AQuESTT Performance Classification Task Force Report

Section I:

AQuESTT Classification Background

In April 2014, the Nebraska Legislature Passed LB438, resulting in Sections 79-760.06 and 79-760.07, which required the Nebraska Department of Education to classify every school and district into a performance level and to identify three schools in the lowest performance category as priority schools. Through the work of a Task Force, the Nebraska Department of Education developed the classification component by which to place every Nebraska school and district in one of four performance levels: Excellent, Great, Good, or Needs Improvement, and to identify the three priority schools.

Originally named the NePAS 1.1 Task Force, the group was renamed to be the Performance Classification Task Force in May of 2014. The group met multiple times over a two year span to establish priorities, develop guiding principles of the model, review research, consider models used in other states, and to participate in the Dominant Profile Judgment method of accountability development, which resulted in several models being developed.

Two classification models were agreed upon during July 2014 Task Force meetings. Afterwards, a smaller Task Force Work Group was formed from members from the larger group. After continued deliberations through November 2014, the Task Force Work Group made a final recommendation to be taken to the State Board of Education.

The Nebraska State Board of Education approved a model for the Performance Classifications System in March of 2015: [AQuESTT Classification Component](#)

In July 2015 a prototype of the AQuESTT Classification Component, now named the AQuESTT Raw Classification, was provided in a secure portal to each district.

The business rules are being reviewed in preparation for the October 2015 release of the AQuESTT Raw Classification.

Section II:

Task Force Members

Bill Auty, Consultant, Nebraska Department of Education
Patty Bentzinger, Norris School Board Member, Norris District 160
Pam Boehle, Curriculum and Assessment Coordinator, Umonohon Nation Public School
Cheryl Burrell, Assistant Superintendent, Winnebago Public Schools
Joe Crum, Data and Assessment Associate – Assessment Research and Evaluation, Millard Public Schools
Dan Dandurand, Technical Coordinator, South Sioux City Community Schools
Julie Downing, Professional development, ESU #13
Leslie Eastman, Assistant Superintendent for Evaluation, Lincoln Public Schools/ESU #18
Dean Folkers, Data, Research and Evaluation, Nebraska Department of Education
Valorie Foy, Director of Assessment and Accountability, Nebraska Department of Education
Kristi Gibbs, Assistant Superintendent for Learning, Ralston Public Schools
Deeann Goeser, Data/Research, Omaha Public Schools
Michaela Goracke, Language Arts Teacher, Sandy Creek Middle School
Cindy Gray, Associate Superintendent, Elkhorn Public Schools
John Hakonson, Superintendent, Lexington Public Schools
Michael Hart, Principal – Norfolk Junior High School, Norfolk Public Schools
Frank Harwood, Superintendent, Bellevue Public Schools
Matt Hastings, Data, Research and Evaluation, Nebraska Department of Education
Sharon Heater, SPED, Nebraska Department of Education
Tracy Heilman, District Assessment Contact, South Sioux City Community Schools
Jeremy Heneger, Assessment and Accountability, Nebraska Department of Education
Renee Hyde, Assistant Superintendent, Papillion – La Vista Public Schools
Kami Jessop, Director of Special Services, Westside Community Schools
Darin Kelberlau, Director of Assessment, Research and Evaluation, Millard Public Schools
Kim Korgan, Data/Research, McCook Public Schools
Freida Lange, Accreditation and School Improvement, Nebraska Department of Education
Mary Ann Losh, Federal Programs, Nebraska Department of Education
Cara Lucas-Richt, SPED Director, Blair Community Schools
Brain Maher, Former Superintendent, Kearney Public Schools
Terri Martin, District Assessment Contact, Gering Public Schools
Russ Masco, Data, Research and Evaluation, Nebraska Department of Education
Kyle McGowen, Superintendent, Crete Public Schools
Dennis McGuire, Accreditation, Nebraska Department of Education
Linda Meyers, Early Childhood Education, Nebraska Department of Education
Dustin Mitchell, Principal – Sutherland High School, Sutherland Public Schools
John Moon, Assessment and Accountability, Nebraska Department of Education
Carla Noerrlinger, Data/Research, Omaha Public Schools
Jennifer Reid, ELL Director, Millard Public Schools – Don Stroh Administration Center
Carol Rempp, Multicultural and Native American, Nebraska Department of Education
Donlynn Rice, Curriculum, Nebraska Department of Education
Sara Robinson, Teacher, Gates Elementary, Grand Island Public Schools
Kris Schneider, ELL Director, Grand Island Public Schools
Terri Schuster, ELL, Nebraska Department of Education
Jay Sears, Program Director, Nebraska State Education Association

Task Force Members - Continued

Amy Shane, Superintendent, O’Neill Public Schools
Gayle Sharkey, District Assessment Contact, North Platte Public Schools
Mike Sieh, Superintendent, Stanton Community Schools
Deila Steiner, Federal Programs Director, Lincoln Public Schools
Diane Stuehmer, Federal Programs, Nebraska Department of Education
John Skretta, Superintendent, Norris District 160
Liz Standish, School Improvement, Lincoln Public Schools
Dr. Bob Uhing, Administrator, ESU #1
Tami Williams, District Assessment Contact, Millard Public Schools
Angie Wright, Curriculum, Elkhorn Public Schools
Brenda Zabel, Teacher, Westside High School

Section III:

Work of the Task Force

***AQuESTT* is: Accountability for a Quality Education System, Today and Tomorrow**

- A framework for a quality education system
- An opportunity to address “accountability” based on Nebraska’s needs

Task Force Membership

- Various size schools
- Varying membership — high percentage at-risk populations to low percentage
- Geographic representation
- Administrative, English Language Learner (ELL) and special education (Sped) program directors, representatives, and teachers
- Districts that represent various demographics: Hispanic, Native American, and Black or African American
- ESU representatives
- Policy Partners/Others
- NDE
- Three national experts
 - Brian Gong — Center for Educational Assessment
 - Chad Buckendahl — Alpine
 - Bill Auty — Educational Measurement — all the psychometric work

Task Force Process

- Reviewed research and related literature
- Reviewed other states’ models and components thereof
- Developed commonality of state perspective on the philosophy of the model and indicators within the model
- Developed Priorities and Guiding Principles
- Reviewed state data
- Made decisions around preliminary indicators

Task Force Priorities

- The Task Force determined the following priorities:
 - Providing/Working within a model that is fair
 - Providing assistance to schools in need
 - Helping all schools to improve
 - **Improving student achievement**

Task Force Guiding Principles

- The Task Force determined the following Guiding Principles:
 - Include Multiple Indicators: Reading, Math, Science, and Writing
 - Incorporate trend data
 - Incorporate Status, Improvement, and Growth
 - Fair and sensitive to change
 - Transparent

Task Force Process Discussed indicators to include

- Components to include:
 - Usable data that is available
 - Consistent data
 - Discriminating data
 - Fair notice to schools and districts
- Vision for other indicators that could be utilized in the future

Task Force Process Continued

- Worked in groups to develop models for accountability, that included standard setting
- Narrowed to two models and ran the data
- Narrowed to one model and ran the data multiple times
- Sent forward as recommendation to State Board of Education

Process – Reasonableness Check

- Could varied sized schools earn a 4, 3, 2, or 1?
- Could schools with high percentages of at-risk students earn high classifications?
- Could varied size schools earn compensatory increases?
- Did the measures used minimize differences in school size as much as possible?
- What unexpected consequences occurred?

Section IV:

Result: Performance Level Classification System (Now Named the Raw Classification)

Each of the following tables provides:

- A decision made by the Performance Classification Task Force
- The process of the Task Force in arriving at the decision
- The rationale for each decision

Decision	<i>N</i> = 25
Process	<ul style="list-style-type: none"> • Reviewed historical NeSA data for stability • Considered school sizes in Nebraska
Rationale	<ul style="list-style-type: none"> • <i>N</i> of 25 showed fairly stable results from year-to-year on NeSA tests • Lower than 25, the increased variability • Higher than 25 eliminated more schools
Business Rule	In any calculation a minimum <i>N</i> of 25 will be applied. If the <i>N</i> is lower than 25, the specific calculation may be dropped out (Growth) or years may be combined (graduation).

Table 1

Decision	To create a model based on a rubric, not an index
Process	<ul style="list-style-type: none"> • Reviewed other states' models • Determined that an index involves high number of points • Determined that points allow ranking of schools
Rationale	<ul style="list-style-type: none"> • Did not want to develop an accountability model that encourages ranking • Rankings are not meaningful differentiation of schools' accountability
Business Rule	Rubric Developed

Table 2

Decision Four Classification Levels	
Process	Reviewed Classifications Levels from 49 states
Rationale	<ul style="list-style-type: none"> • Too many levels could result in insignificant differences between levels. • Five levels could be translated to A-F by the media/public. • Three levels seemed to provide little differentiation. • Four levels allowed for focus on the limited number of schools in the lowest level.
Business Rule	<u>AQuESTT Classification Levels</u> <ul style="list-style-type: none"> • Excellent • Great • Good • Needs Improvement

Table 3

Decision Impact Data — Standard Setting	
Process	<ul style="list-style-type: none"> • Developed 20 models based on various indexes — developed based primarily on modified application of the Dominant Profile Judgment method (Plake, Hambleton, & Jaeger, 1997) • Each model included standard setting
Rationale	<ul style="list-style-type: none"> • Nebraska has quality schools. • Schools identified as “priority” are to come from the lowest performance level; therefore, the number of schools placed in the lowest level needs to allow for focus of efforts.
Business Rule	<u>AQuESTT Classification Levels</u> <ul style="list-style-type: none"> • Excellent = 15% of Nebraska schools • Great = 50% • Good = 30% • Needs Improvement = 5%

Table 4

Decision Include students that are enrolled in a school for the Full Academic Year (FAY)	
Process	<ul style="list-style-type: none"> • Reviewed No Child Left Behind (NCLB) processes • Discussed meaningfulness of accountability as related to opportunity to instruct students
Rationale	If schools are held accountable for student learning through the classification system, schools should have the opportunity to instruct.
Business Rule	Include only FAY students, enrolled from the last Friday in September through the testing window.

Table 5

Decision	Start with Status
Process	<ul style="list-style-type: none"> Reviewed other states' models Discussed meaning of "status" — goal — that students are proficient
Rationale	<ul style="list-style-type: none"> It is important to know that students are on track and will graduate with a certain level of knowledge or skill (proficiency). Allows measure of progress toward goal for achievement at exiting point (graduation) Important to know current level of achievement for each student, building, and district
Business Rule	Status is the average of scale scores of all students and all tests in a school.

Table 6

Decision	Include all scores for a school in one calculation, which naturally weights writing and science less than reading and math
Process	Work Group determined that each test score would count for every student (as long as meet the minimum $N = 25$)
Rationale	<ul style="list-style-type: none"> Wanted all tested subjects to be valued. If included reading, math, science, and writing as $\frac{1}{4}$ of the calculation, large number of schools earned a classification determination based on reading and math only because of the minimum N of 25
Business Rule	Status is the average of scale scores of all students and all tests and is determined by a single calculation. The NeSA-Writing scale score of 70 is linked to scale score of 200.

Table 7

Decision	Improvement is included as a compensatory indicator. Improvement is defined as all scores for NeSA tests two consecutive years (program data).
Process	<ul style="list-style-type: none"> Reviewed Nebraska models Discussed the impact of improvement as included Discussed small school vs large school improvement
Rationale	<ul style="list-style-type: none"> Recognize and reward systematic changes that affect students at a school level Use of trend in improvement was to smooth fluctuations in smaller schools, including using 3 instead of 2 years
Business Rule	Improvement can raise a school's classification level: slope of the three-year line of regression.

Table 8

Growth is included as a compensatory indicator. Growth is a measure of change in score for students enrolled FAY with two years of scores (matched scores in Reading and Math-cohort data).	
Decision	
Process	<ul style="list-style-type: none"> Reviewed Nebraska models Discussed the impact of growth as included Discussed small school vs large school growth
Rationale	<ul style="list-style-type: none"> NeSA is not designed to identify growth at the advanced level (not a vertical scale) Did not want to penalize a school for regression to mean. Designed to reward growth even if performance level does not change Model is fairly transparent, especially when compared to student growth percentiles Growth to apply to all students to avoid attention to only bubble students
Business Rule	Growth can raise a school's classification level based on the Growth Chart (Fig. 1).

Table 9

		Current Year			
Previous Year	1 = Growth 0 = No growth	Exceeds	Met		Not Met
	Exceeds	1	0		0
	Met	1	gain < zero	gain >= zero	0
			0	1	
	Not Met	1	1		gain <= zero
0					1

Figure 1: Growth Chart

Decision Non-Proficient as Supergroup	
Process	<ul style="list-style-type: none"> Reviewed supergroups in other states' models Discussed inclusion of schools — some schools do not have ELL, Sped, or Free and Reduced Lunch to be included Reviewed inclusion of schools with non-proficient
Rationale	<ul style="list-style-type: none"> To include as many schools as possible in this category of the classification system To ensure districts/schools have to focus on any student not proficient regardless of demographic factors Being non-proficient makes a student at-risk
Business Rule	Use only students who score below Meets in calculation for supergroup

Table 10

Decision	<ul style="list-style-type: none"> Increase school classification one level for decreasing non-proficiency percentage beyond the cut score. Decrease school classification one level for increasing non-Proficiency percentage beyond the cut score.
Process	<ul style="list-style-type: none"> Reviewed states' models Reviewed Nebraska models Discussed the need to help all students toward proficiency
Rationale	<ul style="list-style-type: none"> Reward schools/districts for significantly decreasing the number of non-proficient students while holding districts/schools accountable if the number of non-proficient students significantly increases
Business Rule	Non-proficiency change can raise or lower a school's classification level based on the Growth Chart (Fig. 1).

Table 11

Decision		Graduation as ceiling
Process		<ul style="list-style-type: none"> Reviewed Nebraska graduation rates Reviewed national graduation rates Discussed effect of incorporating as compensatory
Rationale		<ul style="list-style-type: none"> Nebraska schools have some of the highest statewide average graduation rates (89.7%) in the nation. The vast majority of Nebraska districts exceed national graduation averages (80%), and numerous districts which are below the state average still eclipse national averages for graduation. Districts should not be penalized for maintaining consistently high graduation rates. Districts should be recognized for meeting certain thresholds of graduation rate (90%/80%). Districts may receive additional recognition for significantly improving graduation rates over time (4 or 7 year cohort, whichever is higher).
Business Rule		<ul style="list-style-type: none"> Based on measure of 4-year or 7-year cohort grad rate, whichever is higher If graduation rate is less than 90%, the classification cannot be Excellent. If graduation rate is less than 80%, the classification cannot be Excellent or Great. If graduation rate is less than 70%, adjust classification to Needs Improvement. N = 25 potential graduates For small schools, multiple years will be used.

Table 12

Decision		Participation
Process		Review of current participation rates: 95% +
Rationale		<ul style="list-style-type: none"> If a school does not test all its students, the validity of school achievement is in question. 95% was selected as the criterion for acceptable participation because that value has been used in state since the implementation of NCLB in 2004. Nebraska schools have been meeting 95%
Business Rule		<ul style="list-style-type: none"> If participation rate is less than 95%, decrease performance level by one. If participation rate is less than 90%, decrease performance level by two. If participation rate is less than 85%, classification level is Needs Improvement.

Table 13

Section V:

Description of Standard Setting for AQuESTT Raw Classification

Chad Buckendahl and Valorie Foy

Introduction:

The Task Force gave the following impact recommendations:

Classification	
Excellent	15%
Great	50%
Good	30%
Needs Improvement	5%

Rationale for Initial Distribution of Schools in the AQuESTT System:

Standard setting involves application of defined policy expectations to a measurement system and then establishing threshold values in that system to align with those expectations. In developing Nebraska’s proposed accountability model, the Task Force engaged in a series of systematic discussions to define levels of performance for each indicator in the model. This process was accomplished by applying a modification of the Dominant Profile Judgment method that relies on subject matter experts evaluating characteristics of the system relative to the defined policy expectations and making recommendations about what levels within and across indicators support interpretation of an overall classification for schools. However, these recommendations cannot occur in a vacuum. A feature of the Dominant Profile Judgment method, as well as all standard setting methods, is inclusion of impact data. Specifically, this means that prior to making final recommendations, panelists should have an opportunity to refine judgments based on the outcomes of their recommendation. As a result, examinee- or test-based standard setting methods are often supplemented with methods to consider the policy tolerance of recommendations. For Nebraska’s accountability model development we incorporated a second method to provide this evidence.

The second method that the Task Force applied when developing their recommendation was a modification of the Hofstee method. In psychometric analysis, the Hofstee method is often used to establish the boundaries of policy tolerance for standards set utilizing other methodologies. For the Task Force, this meant asking them to use the broad policy definitions for each category and discuss the minimum and maximum percentage of schools that would reasonably be classified in each category. As a policy recommendation, they included the following factors about Nebraska’s education system in these judgments:

- Longitudinal participation rates
- Longitudinal 4-year graduation rates
- Statewide performance on external academic measures such as ACT, NAEP

- Intent of the legislation to identify three Priority Schools
- Policymaker interpretation of recommended distribution
- Public interpretation of recommended distribution
- Incentivizing schools to improve on current performance

The result of these discussions yielded an initial distribution of approximately 15%, 50%, 30%, and 5% in the classification categories of Excellent, Great, Good, and Needs Improvement, respectively. These were approximations and not fixed values with intent to establish a baseline from which schools could move from year to year. The final recommendations were a result of lengthy discussions among the Task Force about what constituted a reasonable representation of reality across schools in Nebraska. This representation is ultimately the outcome of any credible standard setting activity.

Example of the Methodology Applied to Graduation

An example of the methodology described above is the deliberation that occurred in determining method of calculating graduation rate in the classification system of AQuESTT.

The Task Force determined that using the calculation for four-year graduation rate or seven-year graduation rate, whichever was higher, would provide opportunity for all districts, including those with a high number of at-risk students, to be compensated in the classification calculations. The calculation would compensate schools for efforts to help at-risk students graduate for the maximum time possible, seven years.

In addition, the Task Force discussed how to include the calculation in the classification. It was first suggested that schools with graduation rates above 90% be given a compensatory increase of one level. However, the Task Force noted that approximately 75% of Nebraska's high schools have graduation rates at or better than 90%. A compensatory indicator that is not discriminating is not an effective calculation to be used in an accountability system. Members then noted, however, that decreasing a district's rating for having less than 90% seemed ineffective due to negative impact for a district that is doing a good job. The Task Force also noted that the graduation rate in Nebraska is the second highest in the nation. For these reasons, the Task Force determined using graduation rate as a limitation or ceiling, rather than as a compensatory indicator.

Example: Graduation-Ceiling

Process of Task Force

- Reviewed Nebraska graduation rates
- Reviewed national graduation rates
- Discussed effect of incorporating as compensatory

Rationale

- Nebraska schools have some of the highest statewide average graduation rates (89.7%) in the nation.

- The vast majority of Nebraska districts exceed national graduation averages (80%), and numerous districts which are below the state average still significantly eclipse national averages for graduation.
- Districts should not be penalized for maintaining consistently high graduation rates.
- Districts should be recognized for meeting certain thresholds of graduation rate (90%/80%).
- Districts may receive additional recognition for significantly improving graduation rates over time (4 or 7 year cohort, whichever is higher).

Business Rule

- Based on measure of 4-year or 7-year cohort graduation rate—whichever is higher.
- If graduation rate is less than 90%, the classification cannot be Excellent.
- If graduation rate is less than 80%, the classification cannot be Excellent or Great.
- If graduation rate is less than 70%, adjust classification to Needs Improvement.
- N=25 potential Graduates
- For small schools multiple years will be used.

Literature Cited

- Hofstee, W. K. B. (1983). The case for compromise in educational selection and grading. In S. B. Anderson and J. S. Helmick (Eds.), *On educational testing* (pp. 109-127). San Francisco, CA: Jossey-Bass.

Section VI:

Summation Statements

- AQuESTT Raw Classification Component is designed as a criterion referenced system.
- AQuESTT Raw Classification Component is a subset of the full AQuESTT accountability system.
- AQuESTT Raw Classification Component is designed to contribute information to school improvement efforts.
- Addition of indicators that can further represent the many efforts of schools is beneficial.
- Addition of indicators must be balanced with providing an opportunity for schools to set goals and manage school improvement to achieve them.

Section VII:

Future Plans

Next

- State Board reviewing final decisions
- Reporting display for State of the Schools Reporting being developed
- More information to come

Business Rules

- John Moon — Available when all indicators are complete