Developing the NePAS: Processes, Outcomes, and Recommendations

Executive Summary

Chad W. Buckendahl, Ph.D.
Alpine Testing Solutions

Bill Auty, Ph.D.
Education Measurement

Brian Gong, Ph.D.
National Center for Improvement of Educational Assessment

July 28, 2014
Acknowledgments

We would like to begin by commending the Nebraska State Board of Education for their thoughtful approach and support in the development of the NePAS system. An accountability system that is designed with input from the range of stakeholders involved in the system has a better opportunity to be viewed as having greater credibility both internally and externally.

In addition, we would like to acknowledge several individuals who assisted us with this accountability system development process. Specifically, Dr. Valorie Foy, Dr. Marilyn Peterson, Tammi Peterson, and many other staff from Nebraska Department of Education were very helpful in designing, planning, and organizing materials and logistics for the study and in supplying data needed to provide feedback to the panelists. The success of the study was due, in large part, to their efforts.

The Task Force of educators representing Nebraska schools and students were the primary contributors to the study as their expertise formed the basis for the recommendations. Specifically, they provided input on the guiding principles for the system, helped to identify indicators of school and district performance, drafted models incorporating those indicators, and revised those models following further discussion and impact data. Without their patience and efforts during the process, there would have been no meaningful outputs.
Executive Summary

The purpose of this project was to develop a range of recommended models for Nebraska’s School Accountability system (i.e., NePAS) that could be communicated to the State Board of Education who would then be tasked with making a final determination. The project involved a series of systematic, reasoned steps to define guiding principles for an accountability model, determine indicators that contribute to interpretations of school and district performance, determine recommended policy statements that characterize the combination of indicators that result in holistic performance, and recommend multiple models that consider the range of factors at the heart of the accountability system. The purpose of this executive summary is to provide a summary of the methods, procedures, analyses, results, and recommendations of the project that was conducted over a series of in-person meetings that occurred in Lincoln, NE, February 24-25 (Meeting 1) March 20-21 (Meeting 2), April 16-17 (Meeting 3), and July 23-24 (Meeting 4), 2014.

The recommended accountability models were developed based primarily on a modified application of the Dominant Profile Judgment method (Plake, Hambleton, & Jaeger, 1997), a standard setting method designed to be used with complex performance assessments. The application of this methodology is analogous to the charge of the Task Force in this project given the multiple indicators, degrees of performance, and the variety of decision rules that could produce an overall classification decision. For this project, the methodology was modified to include development of both indicator level performance and policy statements that defined overall classification decisions driven by the primary purpose of the accountability model: to identify priority schools that would be eligible for identification as potentially needing additional assistance to succeed.

In the subsequent sections of this report, a summary of the activities of each meeting are described to provide context for the recommended models that will be presented to the State Board of Education for consideration.

Meeting 1: Defining guiding principles and indicators

Approximately 50 educators agreed to serve on the NePAS Task Force with representation of classroom teachers, principals, superintendents, professional development, and stakeholder groups. In addition, school size, student population, and geographic location were considered to ensure diversity of perspectives on the Task Force. Most participants were able to attend and contribute to all four meetings. For the first meeting several research documents on accountability were shared with Task Force members. In addition, information on Growth Models, Subgroup/Supergroup, and Minimum-N was presented. The Task Force was asked to develop a set of guiding principles that would be used as a reference point when evaluating draft models in subsequent meetings. Some of these bright line principles included intent that a recommended model would be based on criterion-referenced interpretation of performance that resulted in classification of schools or districts, viewed as fair to all schools regardless of size, transparent, and communicable to diverse audiences of stakeholders.

A secondary task of this first meeting was to begin to determine the indicators that would be included in a recommended model. Although there were many good suggestions for additional indicators that could be included in a future model, the operational needs for the
system required that the Task Force be somewhat limited in the parameters they could consider at this time. Specifically, indicators were limited by data that are currently available in the state’s student information management system or could be calculated using these data. As this system evolves, so, too, can the NePAS. Based on the current available information, the following indicators were included in models that the Task Force considered: Status for Reading, Mathematics, Science, and Writing; Improvement for Reading, Mathematics, Science, and Writing; Growth for Reading and Mathematics; Graduation Rate; and Participation Rate. Defining and combining these indicators to determine an overall district classification occurred during the second meeting. The committee reviewed the naming of performance levels.

Meeting 2: Additional research documents were provided to Task Force members. Members refined the discussion of priorities and Guiding Principles and received further research on use of subgroups and supergroups. Information was presented on measurement of simple growth, student growth percentiles, and adequate student growth—and the Task Force reviewed two models from other states more extensively: Massachusetts and Idaho. The group established a philosophy of basing the system on a low number of points in the accountability system so schools could not be ranked from 1 to 966, and determined that a second pass would be established for identifying the “priority schools” from among schools that fell into the lowest performance level. Graduation rate was discussed and the Task Force determined that using four-year and seven-year graduation rates best represented the efforts of districts to meet the needs of all students. Also reviewed were federal accountability options, reduction of achievement gaps, goal of 100% proficient by 2020, and the ESEA waiver application process.

Meeting 3: Drafting accountability models and policy statements

The primary tasks for the third meeting involved workshop activities that were designed to define the indicators of performance along with policy statements that would yield overall school classification ratings. To accomplish this goal, the Task Force was subdivided into 10 tables of approximately 4-5 participants per table with intentional assignment to ensure diversity of stakeholder group representation at the table. Each group’s charge was to independently propose a model for two grade configurations defining the indicators determined at the first meeting and then combining those indicators to yield an overall classification for a school. Four grade level configurations representing the most common found across the state were assigned. Specifically, elementary (K-5), middle (6-8), high school (9-12), and “all grades” (K-12) configurations were considered by the group. Because the meeting occurred over two days, this meant that 5 groups were assigned to each possible configuration to develop, in concept, 20 models that would apply to the respective grade level configuration. Following the meeting, facilitators engaged in a series of analyses to evaluate and synthesize the models before calculating the potential impact of each prior to the next meeting.

Although 20 models were submitted based on the charge, the groups’ models were effectively the same for their first and second assigned grade configuration, permitting an initial reduction of considerations to 10. Of these 10 remaining models, there was clear overlap of concepts among many of the groups suggesting that these were not unique. Specifically, at the indicator level, the common concepts that appeared among all proposed models involved criterion-referenced interpretations, norm-referenced interpretations, or a combination of these types of interpretations. At the policy recommendation level that described profiles of schools at
a given classification level (generically characterized as levels 1, 2, 3, and 4 to avoid placing value-laden labels at this stage of the process), groups suggested compensatory, conjunctive, or a combination of these decision rules. Although groups may have used different indicator level threshold values, the core philosophies were characterized in the three models that were further analyzed and carried forward to the third meeting.

**Meeting 4: Evaluating impact data and refining draft models**

At the fourth meeting, Task Force members were asked to review the synthesized models, evaluate them against their guiding principles, consider impact data calculated for the three proposed models, consider additional policy considerations, and recommend revisions to the remaining models that could be communicated to the State Board of Education for consideration. In reviewing the synthesized models, panelists were asked make an initial determination of the viability of each relative to the guiding principles. Following this triage activity, one of the models was rejected as being potentially too difficult to communicate to a broad group of stakeholders and as violating a goal of maintaining an overall classification based on a rating profile.

The Task Force then spent the remaining time during the final meeting revising the two models that they deemed viable and consistent with the guiding principles. For this activity, the group was subdivided into six tables with approximately 4-5 participants per table with assignment to a table being intentional to ensure diverse representation. Part of the policy considerations that the group included in their deliberations was reality check that the impact data provided as well as the purpose of the system to identify priority schools that would potentially be eligible for assistance. Impact data were brought into the process only at the last meeting to ensure that the process and judgments were anchored in the expectations of the indicators with consideration of these data as an informed check on the external credibility of the recommendations.

Following independent work by each table to suggest a policy philosophy for each model, determine how subgroup information would be considered, determine how growth would be considered, and any additional revisions to the proposed model, three tables assigned to each model further consolidated their discussion to yield two models based on similar philosophies, but with some differences in terms of how indicators are defined and how these indicators are combined to determine an overall classification for a school or district. The results of these activities are the two models recommended for consideration by the State Board of Education. The Task Force completed their work with a procedural evaluation form that solicited information about individual participants’ confidence in the training activities, the model development and revision processes, and the final recommendations.
Resources Reviewed by Task Force Members


References