

Performance Standard Setting for Kentucky's School Accountability System: Overview for 2018 and 2019

Brian Gong and Chris Domaleski

Center for Assessment

**Presentation for discussion with Kentucky Board of Education
June 6, 2018 Frankfort, KY**

Overview of presentation

- Background and context of accountability performance standard-setting
- Description of accountability system standard-setting for 2018
- Preview of accountability system standard-setting for 2019
- Discussion and identification of any needed follow-up

Accountability system purposes

Purposes of Kentucky's school accountability system include:

- Provide useful information regarding school performance on important measures to help direct improvement efforts, foster student proficiency, and close achievement gaps
- Identify schools in need of mandated support from districts and Kentucky Department of Education
- Comply with state and federal law

Accountability system background

- Kentucky has had a school accountability system since 1992
- Most recent proposal complies with Kentucky Senate Bill 1 (SB1) (2017, General Assembly) and federal Every Student Succeeds Act (ESSA) (received full approval 5/7/18)
- Notes:
 - SB 1 prohibits a single numeric score
 - ESSA requires identification of schools for Comprehensive and Targeted Support and Improvement (CSI/TSI) starting fall 2018
 - Kentucky approved to do focused system in 2018, and full 5-star system starting 2019 when more data elements will be available
 - Kentucky plans to report additional data that do not affect accountability ratings
- School accountability system interacts with many other programs intended to promote quality, including graduation requirements; district and state support systems; curriculum and instructional programs such as content standards, CTE; assessment and reporting systems, etc.

Accountability system description

- 2018 – focused system
 - Identify CSI/TSI schools
 - Four federal **indicators** at elementary/middle and high school, each with associated **measures**; each measure has a **metric** for accountability
 - For example: Proficiency Indicator includes reading and mathematics
 - Student metric: Novice/Apprentice/Proficient/Distinguished (NAPD) achievement level based on performance on state assessment
 - School accountability metric: weighted average of student performance for reading and math *combined* for all students for whom school is accountable, by assigning points to achievement levels, e.g., N=0, A=.50, P=1.00, D=1.25

Accountability standard-setting: general

- Accountability systems typically define:
 - What is measured
 - How much is “good enough”
 - Consequences for (not) meeting “good enough” criteria
- The process for establishing what is “good enough” is called performance “standard setting”
- Standard setting has a long tradition for establishing assessment proficiency/achievement levels, e.g., NAPD
 - Systematic, fair/representative, transparent, documented
- Kentucky is among states leading the way using a systematic standard-setting process for establishing performance criteria for accountability systems
 - More challenging than assessment standard-setting because accountability systems include more performance indicators that can combine in complex ways and schools may be quite varied (e.g., size, student demographics, grade configuration)

Accountability standard setting – 2018

- Outcome: Be able to assign schools to CSI/TSI/Other
- Constraints
 - CSI is “Title I or non-Title I school performing as poorly as bottom 5% of Title I schools”
 - TSI is “Title I or non-Title I school with one or more student groups performing as poorly as all students in the lowest performing 5% Title I schools [by elementary/middle/high school level]” (Consolidated State Plan, p. 85)
 - “Other” is all other schools not identified as CSI or TSI
 - Because of SB1, need to devise a way to identify bottom 5% of Title I schools without having an overall numeric rating
 - Want to identify bottom 5% of Title I schools for CSI but not identify many more. Total identified likely to be somewhat greater than 5%

Proposed method

- General approach: Use “school profiles” of performance, with empirical data, to identify cutpoint for bottom 5% of Title I schools
- Specific approach: Apply successively ordered school profiles to identify schools; stop when target amount of schools have been identified
- Standard-setting: a) confirm weights for multiple measures within indicators, b) establish the profiles of indicators, c) establish the order of profiles; d) check that profiles work as intended
 - Must comply with priorities identified by the Kentucky Board of Education
 - Must comply with requirements of ESSA (e.g., School Quality/Student Success (SQSS) must count less than academic indicators together)
 - Should identify 5% of Title I schools consistent with Performance Level Descriptors (PLDs)

Example of method

- Example sequence of profiles of performance for identifying bottom 5% of Title 1 schools, high schools (Consolidated State Plan, p. 83)

| Profile | Graduation Rate | Academic Ach. | ELP Progress | Sch. Quality Stu. Success |
|---------|-----------------|---------------|--------------|---------------------------|
| 1 | Bottom 5% | Bottom 5% | Bottom 5% | Bottom 5% |
| 2 | Bottom 10% | Bottom 5% | Bottom 5% | Bottom 10% |
| 3 | Bottom 10% | Bottom 10% | Bottom 10% | Bottom 10% |
| 4 | Bottom 15% | Bottom 10% | Bottom 10% | Bottom 15% |

Notes about 2018 standard setting

- Task is to create small gradations of *less than 5%* of Title I schools at the very low end of performance
 - Setting CSI identification for 2018 is sufficient for identifying CSI Title I and non-Title I schools, and for identifying TSI schools
 - Very different and much less complex task than 5-star standard setting, which will rate performance across the full spectrum of school performance
- 2018 accountability system standard setting logistics
 - Held late August (earliest data availability)
 - About 15 panelists, with heavy representation of policy-oriented persons; ideal to have some panelists who know the priorities established by the Kentucky Board of Education in Kentucky's Consolidated State Plan (CSP)
 - Special Kentucky Board of Education meeting early September to review and approve Commissioner's recommendations from the standard setting work
 - Accountability standard-setting results applied to school identification of CSI, TSI and Other (mid-September)

Accountability standard setting – 2019

Future Planning for Next Year

- Outcome: Be able to assign all schools to 5-star ratings, as well as to CSI and TSI
- 5-Star ratings determined by profile plus special decision rules (e.g., “No higher than 3 Stars if...”)
- Will have additional data elements to include in ratings (and addition for reporting only)

Proposed method – 2019

- Set performance standards on individual indicators that combine multiple measures (e.g., “Very Low” to “Very High” on Proficiency)
- Set performance standards on profiles (combinations) of indicators
- Set performance standards of star ratings including special decision rules (e.g., “No higher than 3 Stars if...”)
- Set performance standards separately for elementary, middle and high schools; check for coherence across grade spans

Comparing decision matrix and index approaches

- Index approach is very good for combining multiple different performance measures because it provides a common scale, and provides a straightforward way to assign nominal weights
- Decision matrix approach is very good for combining multiple different performance measures when non-linear or irregular relationships are needed/desired, or when an overall rating is desired but not an overall scaled score
 - Avoids some problems with highly compensatory systems, effective weights

Example of multiple weights for a single measure

- Sometimes it is not desirable to have a single weight for a measure in an accountability system
- For example at elementary and middle schools, the system includes indicators of proficiency and growth

| Average Proficiency Level of School | Weight for Proficiency | Weight for Growth |
|-------------------------------------|------------------------|-------------------|
| High | 90% | 10% |
| Medium-high | 75 | 25 |
| Medium-low | 50 | 50 |
| Low | 10 | 90 |

Example of a decision matrix

- A decision matrix can reflect different emphases or values (like weights) in non-linear ways, especially with categorical variables

Sample decision matrix, two variables each with 5 levels (1=lowest; 5=highest); overall rating has 5 levels (VL to VH)

| | | Variable 2 | | | | |
|------------|---|------------|----|---|---|----|
| | | 1 | 2 | 3 | 4 | 5 |
| Variable 1 | 5 | M | M | H | H | VH |
| | 4 | M | M | M | H | VH |
| | 3 | L | L | M | H | H |
| | 2 | VL | L | L | M | M |
| | 1 | VL | VL | L | L | L |

Notes about 2019 standard setting

- Task is to produce a coherent and manageable mapping between 5 Stars and a very large set of possible combinations (profiles) of indicators, some of which may have multiple measures combined in complex ways; need to consider reliability
- Standard setting process must be simple enough panelists can do it well; will require structure from organizers
- Will need to have clean data so that it can inform standard setting with empirical information
- Similar logistics are envisioned as 2018 standard setting; if possible, move earlier in summer for 2019

Discussion, identify needed follow-up

For more information:

Center for Assessment
www.nciea.org



Chris Domaleski

cdomaleski@nciea.org

Brian Gong

bgong@nciea.org