# Accountability System Development for 2013 and Beyond
# Framework Options and Concerns

## Introduction

The intent of the accountability development process is to design a new accountability system aligned with the provisions of House Bill (HB) 3 (81st Texas Legislature, Regular Session). Every aspect of the accountability system will be reevaluated. The resulting accountability system may look very different from the current state accountability system. A defining characteristic of the new accountability system will be the emphasis on postsecondary-ready performance on the new State of Texas Assessments of Academic Readiness (STAAR).

The framework of the accountability system describes the way performance on the accountability indicators is evaluated for ratings. The choice of a framework will guide the accountability development process. Two types of frameworks meet the accountability requirements in HB 3 – Separate Indicators and Performance Index.

## Separate Indicators

A Separate Indicators framework requires districts and campuses to meet accountability targets on each performance measure. This framework clearly identifies the lowest-performing subject, student group, or other indicator. The former state and federal accountability systems were based on a Separate Indicators framework. However, while performance was evaluated separately for each subject and student group, these systems are considered modified Separate Indicators systems because they included other features that provided campuses and districts with alternatives to meeting absolute targets. For example, both systems allowed campuses and districts that did not meet absolute accountability targets to meet other criteria that were considered equivalent to the absolute targets (e.g., required improvement). Additionally, the state system included an exceptions provision that allowed campuses and districts to meet the targets on a proportion of the performance measures as long as they did not repeatedly fail on the same measures. These systems also incorporated both absolute performance and student progress. Through these features, campuses and districts not meeting absolute targets were able to receive higher ratings than would be available to them in a pure Separate Indicators framework.

Under the state accountability system in 2011, *Academically Acceptable* ratings required campuses and districts to meet criteria on up to 25 separate assessment measures (five subjects times five student groups), plus up to 10 dropout and high school completion measures. HB 3 expands the scope of the accountability system and increases the number of indicators that must be used. One option for the new state accountability system is to preserve this framework, but the increase in number of indicators will require trade-offs.

## Performance Index

With a Performance Index each measure contributes points to an index score. Districts and campuses are required to meet one accountability target on the total index score. Performance on all measures is included, but stronger performance in some areas can compensate for weaker performance in other areas. With a Performance Index, the resulting rating reflects overall performance rather than the weakest areas. Many variations are possible such as weighting the measures to reflect state goals. Combining performance into an index would be a new model for Texas. A strength of a Performance Index design is that any number of indicators and student groups can be added to the system without creating additional state targets for campuses and districts to meet.

## Variations

A summary of the key features of the Separate Indicators framework and Performance Index framework is provided in the table that follows. The remainder of the document illustrates the two frameworks, with examples of variations. In general, variations are used to address concerns or weaknesses in the frameworks. Use of a variation can compromise to some extent the greatest strengths of the framework. Some of the variations produce what could be considered a hybrid of the two frameworks.

## Comparison of Separate Indicators and Performance Index System Frameworks

| Separate Indicators | Performance Index |
|---|---|
| Performance is calculated and evaluated separately for each discrete measure and student group included in the system. | Performance is calculated separately for each discrete measure and student group included in the system, but the results of each measure contribute points to a Performance Index. The Performance Index score is evaluated. |
| Performance on each measure must meet a state target. | Overall performance on the index must meet a single state target. |
| Assigns rating based on the lowest performing measure regardless of how well the campus or district performs in other areas. | Assigns rating based on overall campus or district performance regardless of how poorly a campus or district performs in a particular area. |
| Different accountability targets address stronger and weaker performance across subjects, but typically all student groups must meet the same target for the subject. | The overall index score allows stronger performance in one area (subject, student group, or performance level) to compensate to some extent for weaker performance in other areas. |
| Accountability rating and interventions focus attention on campus or district problem areas. | Accountability rating may not focus attention on campus district problem areas, but interventions will. |
| The number of targets that must be met varies from campus to campus based on size and diversity of student population. | All campuses and districts must meet one target. The maximum number of points that make up the index varies from campus to campus based on size and diversity of student population. |
| All measures are weighted equally because performance on any one measure can result in a lower rating. | All measures are not weighted equally because low performance on any one measure will not necessarily result in a lower rating. |
| The number of separate measures will depend on how many measures and student groups are included in the system. | Adding measures to the index does not increase the number of targets that must be met. |
| It may be necessary to set initial state targets low enough to avoid rating excessive numbers of campuses or districts as unacceptable because they must meet the state target on every measure every year. | A higher initial state target could be set because it is not necessary for campuses and districts to meet the target on every measure every year.. |
| A Separate Indicators system can be used to categorize campuses and districts as acceptable or unacceptable but does not produce a ranking. | A Performance Index system can be used to categorize campuses and districts as acceptable or unacceptable. It can also be used to produce a ranking. |
| Federal statute and rules currently require a Separate Indicators system for adequate yearly progress (AYP). | Most states that have integrated a state accountability system with AYP use a Performance Index framework for the state system. |
| Additional features offer alternative ways to meet ratings criteria for each indicator:<br>• Current year performance OR<br>• Three-year average performance OR<br>• Required improvement | Additional features offer alternative ways to meet ratings criteria for the overall index score:<br>• Current year performance on index OR<br>• Three-year average performance on index OR<br>• Required improvement on index |
| The 85 percent proportional feature, which requires campuses and districts to meet state targets on 85 percent of measures on which they are evaluated each year, provides some relief for large, diverse campuses and districts that are evaluated on more state targets. | The 85 percent proportional feature is not applicable to a Performance Index. The index model, by definition, provides relief to large, diverse campuses and districts that are evaluated on many measures. |

### Illustrations

In the following illustrations assume a middle school with grades 6-8 is being evaluated. STAAR Level II performance is evaluated for five subjects and Level III performance is evaluated for mathematics, reading, and writing. The illustrations use the five student groups from the previous state accountability system. Assume no students at this school take any End-of-Course (EOC) assessments. Also, for simplicity, student progress is not addressed nor are graduation rates, dropout rates or completion rates. Options for incorporating these additional indicators will be discussed later.

### Separate Indicators

The table below depicts a Separate Indicators system. The percentages in the cells represent the percent of students meeting the performance level standard, *i.e.*, 60% of all students tested in mathematics met the Level II standard. Eighteen percent of all students tested in mathematics met the Level III standard.

| Indicators | | Student Groups | | | | |
|---|---|---|---|---|---|---|
| | | **All** | **AA** | **H** | **W** | **ED** |
| Mathematics (gr. 6, 7, 8) | Level II | 60% | 59% | 56% | 69% | 54% |
| | Level III | 18% | 15% | 12% | 29% | 10% |
| Reading (gr. 6, 7, 8) | Level II | 68% | 63% | 64% | 74% | 62% |
| | Level III | 21% | 14% | 11% | 25% | 11% |
| Writing (gr. 7) | Level II | 72% | 69% | 71% | 74% | 69% |
| | Level III | 19% | 15% | 16% | 22% | 11% |
| Science (gr. 8) | | 58% | 56% | 50% | 69% | 48% |
| Social Studies (gr. 8) | | 73% | 70% | 70% | 77% | 69% |

*AA—African American; H—Hispanic; W—White; ED—Economically Disadvantaged*

- An acceptable rating requires meeting the target on every measure (40 separate measures in this example assuming every student group met minimum size criteria). If the acceptable target is 50% for Level II for all groups and every subject, and 10% for Level III, this campus would receive an unacceptable rating due to economically disadvantaged student group performance in science.
- The overall rating is based on the lowest performing measure. In this way, the rating targets the problem area. Campuses and districts with smaller populations of economically disadvantaged and ethnic minority students must ensure those student groups meet performance targets.
- Minimum size criteria for evaluation of student group performance require larger campuses and districts with more diverse student populations to meet accountability targets on more measures. All campuses and districts are not evaluated on the same number of measures.
- Every measure receives the same weight in determining the rating – science counts the same as mathematics even though fewer grades are tested in science; any student group that meets minimum size criteria counts the same as All Students.

- Adding indicators to the system increases the number of accountability targets campuses and districts must meet.
- Accountability targets must be set low enough to avoid rating excessive numbers of campuses and districts as unacceptable because they must meet the targets on every subject every year.
- Campuses and districts can be categorized as acceptable/unacceptable but cannot be ranked from highest to lowest performing.
- Setting annual accountability targets assumes a stable assessment program and stable indicator definitions. Without stability a change in a rating could reflect a change in the system rather than a change in academic performance of students at the campus or district.

**Separate Indicators Variations**

Concern 1: Rating is driven by lowest performing group.

Ways to Address:

- Use additional features to allow for ways other than meeting the absolute target to earn a rating. If performance does not meet the absolute target, then allow any or all of the following to help.
  - Meet the target with student progress (credit for students who do not meet the standard but are progressing toward the standard).
  - Meet the target with three-year average performance.
  - Meet a Required Improvement target (campus improvement from prior year).
  - Apply an 85% provision which allows 15% of measures to fall below targets, with safeguards to limit its use.

Concern 2: Large and diverse campuses and districts are accountable for many more measures than smaller, more homogenous campuses and districts, putting them at a disadvantage in ratings.

Ways to Address:

- Use minimum size criteria so that only groups of sufficient size are evaluated.
- Limit the number of student groups evaluated:
  - Limit the number of student groups evaluated for any indicator. Only evaluate performance of the two largest race/ethnicity student groups on the campus/district for any indicator, for example. This results in a maximum of two race/ethnicity groups per indicator for every campus/district.
  - Limit the indicators for which student group performance is evaluated. Evaluate performance of all student groups for reading and mathematics, for example, but evaluate only All Students and Economically Disadvantaged performance for writing, science, and social studies.

Concern 3: Under new statute, there are more indicators to consider than under the previous system. This is largely due to the requirement that both Level II and Level III performance be evaluated as well as progress to Level II and progress to Level III. Also, due to differences in the assessments, it is likely the EOC indicators will be separate from the grade 3-8 indicators. The EOC indicator is likely to be a separate longitudinal, cumulative calculation for each grade cohort. Cohorts will phase in as the system matures. Each of these factors is multiplied by the number of student groups selected as shown in the following table. Although most campuses would be evaluated on only the grade 3-8 indicators or the EOC cohort indicators, about ten percent of campuses and most districts have grade spans that include grade 3-8 and high school.

Note that in the table on the following page only EOC cohorts for grades 9 and 10 are shown. Also, for illustration purposes, the five student groups evaluated under the former system are used, though the number of student groups could be fewer, or more, and could vary by indicator. Student progress measures are not shown.

Ways to Address:

- Phase in the evaluation of student group performance.
- Combine Level II and Level III performance into a single weighted measure. For example, a student achieving the Level III standard could contribute 1.2 points, meeting Level II only would contribute 1.0 points. Progress could be incorporated such that a failing student showing progress could contribute 0.5 points.
- Similarly, combine the high school cohorts into a single weighted EOC indicator for each subject rather than separate indicators for each cohort. The indicator would be weighted according to the size of each cohort because students in higher grades have accumulated more points toward the EOC cumulative score graduation requirement.
- Evaluate All Students performance for each subject individually but sum results across subjects for a single measure of assessment performance for each student group. This would result in the following twelve grade 3-8 STAAR measures, a reduction from the 40 measures in the example. There are different ways to combine performance across subjects but all result in something that reflects average performance across the subjects.
    1. Reading Performance Level II (All Students only)
    2. Reading Performance Level III (All Students only)
    3. Mathematics Performance Level II (All Students only)
    4. Mathematics Performance Level III (All Students only)
    5. Writing Performance Level II (All Students only)
    6. Writing Performance Level III (All Students only)
    7. Science Performance (All Students only)
    8. Social Studies Performance (All Students only)
    9. African American Student Performance (summed across subjects)
    10. Hispanic Student Performance (summed across subjects)
    11. White Student Performance (summed across subjects)
    12. Economically Disadvantaged Student Performance (summed across subjects)

Note that if student group performance is collapsed across subjects, then effectively, the result is a system that is a hybrid of Separate Indicators and Performance Index.

| Possible Indicators (number of measures) | Student Groups* | Total Number of Measures for Row | Total Number of Measures for Type of Indicator |
|---|---|---|---|
| Gr. 3-8 Reading Levels II & III (2) | All, AA, H, W, ED (5) | 10 | 40 |
| Gr. 3-8 Mathematics Levels II & III (2) | All, AA, H, W, ED (5) | 10 | |
| Gr. 4 & 7 Writing Levels II & III (2) | All, AA, H, W, ED (5) | 10 | |
| Gr. 5 & 8 Science Level II (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 8 Social Studies Level II (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 9 EOC Reading (1) | All, AA, H, W, ED (5) | 5 | 25 |
| Gr. 9 EOC Mathematics (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 9 EOC Writing (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 9 EOC Science (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 9 EOC Social Studies (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 10 EOC Reading (1) | All, AA, H, W, ED (5) | 5 | 25 |
| Gr. 10 EOC Mathematics (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 10 EOC Writing (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 10 EOC Science (1) | All, AA, H, W, ED (5) | 5 | |
| Gr. 10 EOC Social Studies (1) | All, AA, H, W, ED (5) | 5 | |
| 4-yr or 5-yr Graduation Rate | All, AA, H, W, ED (5) | 5 | 5 |
| Annual Dropout Rate | All, AA, H, W, ED (5) | 5 | 5 |
| Total Possible | | | 100 |

* *Student groups can be as few as one (All only) or as many as 12 (All, AA, H, W, ED, non-ED, Amer. Ind., Asian, Pac. Islander, Two or More, LEP, Spec. Education.)*

**Performance Index**

The table below depicts a Performance Index system. As in the previous illustration, percentages in the cells represent the percent of students meeting the performance level standard. The performance shown in this illustration is exactly the same as that provided for the Separate Indicators example. In this example, the percentages of students meeting the standards are converted directly to points that are summed to obtain the total score. A 59% meeting Level II in mathematics for African American students contributes 59 points.

| Indicators | | Student Groups | | | | | Index Points | Maximum Index Points |
|---|---|---|---|---|---|---|---|---|
| | | All | AA | H | W | ED | | |
| Mathematics (gr. 6, 7, 8) | Level II | 60% | 59% | 56% | 69% | 54% | 298 | 500 |
| | Level III | 18% | 15% | 12% | 29% | 10% | 84 | 500 |
| Reading (gr. 6, 7, 8) | Level II | 68% | 63% | 64% | 74% | 62% | 331 | 500 |
| | Level III | 21% | 14% | 11% | 25% | 11% | 82 | 500 |
| Writing (gr. 7) | Level II | 72% | 69% | 71% | 74% | 69% | 355 | 500 |
| | Level III | 19% | 15% | 16% | 22% | 11% | 83 | 500 |
| Science (gr. 8) | | 58% | 56% | 50% | 69% | 48% | 281 | 500 |
| Social Studies (gr. 8) | | 73% | 70% | 70% | 77% | 69% | 359 | 500 |
| Total Index Score out of 4000 maximum points | | | | | | | 1873 | 4000 |
| Percent of Maximum Points | | | | | | | 47% | |

*AA—African American; H—Hispanic; W—White; ED—Economically Disadvantaged*

- In this system, an acceptable rating requires meeting a target on one measure – the Percent of Maximum Points of the total index score. (Note: A percentage is calculated since the total number of maximum points can vary from campus to campus.)
- The rating based on overall index score allows stronger performance in one area (subject, student group, or performance level) to compensate to some extent for weaker performance in other areas.
- In this example, the maximum number of points possible is 4000 but this particular number is not required. The points possible depend on the measures included and their maximum point value. Student group measures (as used in this example) will result in differing numbers of maximum index points depending on the campus demographics.
- Additional features can be used to offer other ways to meet the Performance Index target:
  - Meet the target with three-year average performance on the total index score; or,
  - Meet Required Improvement target on change from prior year in total index score.

  The 85% provision is not applicable since the system requires meeting only one target.
- In this example, each subject, student group, and performance level is given the same weight (one percentage point equals one point toward the total score). A variation of this model could be to assign weights to each subject, group, or level as a way to emphasize the relative importance of each cell. This would add complexity to the index.

- Indicators can be added to the system without increasing the number of accountability targets campuses and districts must meet.
- Campuses and districts can be *rated* based on meeting a target for the Percent of Maximum Points in the total index score. If the target is 45%, for example, this campus would receive an acceptable rating even though performance is below 45% on some of the component measures.
- Campuses and districts can also be *ranked* from highest to lowest on Percent of Maximum Points. The ranking could be used to assign ratings – the bottom ten percent of campuses receive an unacceptable rating, for example. Annual accountability ratings based on ranking may be less subject to unintended results during the transition to a stable assessment program and stable indicator definitions.

**Performance Index Variations**

Concern 1: A Performance Index framework is a black box that is difficult to understand or interpret. A single index with a single outcome may be perceived to provide little insight into the strengths and weaknesses of a campus or district.

Ways to Address:

- Create a reporting system that clearly shows each component of the index and the calculations used to determine the total score.
- Rather than creating a single index that includes all measures, create separate indexes for:
  - assessment results and dropout or graduate results, or
  - each subject, or
  - each student group, or
  - each performance level (Level II and Level III)

  Note that if separate indexes are created and each one has to meet a target, then effectively, the result is a system that is a hybrid of Separate Indicators and Performance Index.

The following tables illustrate separate indexes by subject versus separate indexes by student group.

**Separate Subject Indexes**

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| All Students<br>African American<br>American Indian<br>Asian<br>Hispanic<br>Pacific Islander<br>Two or more races<br>White<br>Economically Disadvantaged<br>Non-economically Disadvantaged<br>Limited-English Proficient (LEP)<br>Special Education | Reading Performance (Gr. 3-8, EOC, Levels II and III) | Mathematics Performance (Gr. 3-8, EOC, Levels II and III) | Writing Performance (Gr. 4,7, EOC, Levels II and III) | Science Performance (Gr. 5,8, EOC, Level II) | Social Studies Performance (Gr. 8, EOC, Level II) |
| | Reading Score | Mathematics Score | Writing Score | Science Score | Social Studies Score |

With the Separate Subject Indexes:

- Targets would be set for each *subject*;
- The number of student groups included can be maximized without creating additional targets;
- Poor performance in one student group or one performance level can be offset somewhat by high performance of other groups (as a safeguard to this, minimum performance floors could be set to control if a student group contributes to the subject score);
- Performance gaps between *subjects* are the focus.

**Separate Student Group Indexes**

| Gr. 3-8: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reading (Lvl II) | All Students | African American | American Indian | Asian | Hispanic | Pacific Islander | Two or More Races | White | Economically Disadvantaged. | Non-Economically Disadvantaged. | Limited English Proficient | Special Education |
| Reading (Lvl III) | | | | | | | | | | | | |
| Math (Lvl II) | | | | | | | | | | | | |
| Math (Lvl III) | | | | | | | | | | | | |
| Writing (Lvl II) | | | | | | | | | | | | |
| Writing (Lvl III) | | | | | | | | | | | | |
| Science (Lvl II) | | | | | | | | | | | | |
| Soc. St. (Lvl II) | | | | | | | | | | | | |
| EOC: | | | | | | | | | | | | |
| Reading (Gr. 9) | | | | | | | | | | | | |
| Reading (Gr. 10) | | | | | | | | | | | | |
| Reading (Gr. 11) | | | | | | | | | | | | |
| Math (Gr. 9) | | | | | | | | | | | | |
| Math (Gr. 10) | | | | | | | | | | | | |
| Math (Gr. 11) | | | | | | | | | | | | |
| Writing (Gr. 9) | | | | | | | | | | | | |
| Writing (Gr. 10) | | | | | | | | | | | | |
| Writing (Gr. 11) | | | | | | | | | | | | |
| Science (Gr. 9) | | | | | | | | | | | | |
| Science (Gr. 10) | | | | | | | | | | | | |
| Science (Gr. 11) | | | | | | | | | | | | |
| Soc. St. (Gr. 9) | | | | | | | | | | | | |
| Soc. St. (Gr. 10) | | | | | | | | | | | | |
| Soc. St. (Gr. 11) | | | | | | | | | | | | |
| | All Score | African Amer. Score | Amer. Ind. Score | Asian Score | Hisp. Score | Pac. Is. Score | Two or More Score | White Score | Econ. Score | Non-econ. Score | LEP Score | Spec. Ed. Score |

With the Separate Student Group indexes:

- Targets would be set for each *student group* (the number of student groups would likely be limited from the 12 shown above);
- The number of campuses and districts meeting minimum size criteria for a student group might increase, depending on how performance is combined across subjects;
- The number of subject indicators included can be maximized without creating additional hurdles;
- Poor performance in one subject or at one performance level can be offset somewhat by high performance in other areas (as a safeguard to this, minimum performance floors could be set to control if a subject contributes to the student group score);
- Performance gaps between *student groups* are the focus.

Concern 2: The compensatory nature of a Performance Index system allows for unacceptable performance of one student group, one subject, or one performance level to be masked.

Ways to Address:

- o Include progress measures in the index to reward efforts to close gaps.
- o Assign minimum performance floors for individual measures. Floors could be set statistically, at the 10$^{th}$ percentile of current year performance, for example. Performance below a floor would not contribute any points to the score, which would direct attention to the poorest areas.
- o Assign weights to the individual measures included in the index to place appropriate emphasis on areas of most concern. For example, assign higher weight to progress measures to ensure failing students are a priority or higher weights to Level III to focus on postsecondary-ready performance.

**Use of Performance Index Systems in Other States**

Twenty-five states were profiled in 2010 when developing the HB 3 Transition Plan. Six states, including California, Florida, and North Carolina, combine performance across subjects in a Performance Index. Several states include performance on indicators other than test results in the index. Information about Performance Index accountability systems in several states will be provided at the meeting.