# Draft Recommendations for a New State Accountability Index

State Board of Education
System Performance Accountability

September 24, 2008

Pete Bylsma, EdD, MPA

## TABLE OF CONTENTS


| | |
|---|---|
| Background | 2 |
| Accountability Index | 3 |
| District Accountability | 7 |
| Identifying Priority Schools and Districts (Lowest Tier) | 10 |
| Integrating the Systems | 12 |
| Recognition | 13 |
| | |
| Appendix A – Indicators and Outcomes | 19 |
| Appendix B – Identifying Priority Schools and Districts | 33 |
| Appendix C – Current State Assistance Program | 36 |
| Appendix D – Advisory Group Members | 37 |

## BACKGROUND

The legislature requires the State Board of Education to develop a statewide accountability system. Part of that requirement is to identify schools and districts for recognition and for receiving additional state support. To meet this requirement, the Board is developing an accountability index to sort schools and districts into different "tiers" based on multiple factors. The Board hired Pete Bylsma, an independent consultant, to work with a set of advisors to develop the proposed index for Board review and to identify data that could be used that would help identify "Priority" schools and districts in most need. (Mass Insight is designing a system to support the schools and districts in most need, and this system will be aligned with the system of support that OSPI offers.) This document provides the initial recommendations for the index and information about identifying Priority schools and districts. The Board plans to present a proposal to the 2009 Legislature.

A number of principles guided the development of the system. These include the principles the Board adopted in previous meetings (in bold) and others that reflect feedback about the system and advice from the advisors. Specifically, the accountability system should:

- **Be transparent and simple to understand**;
- **Use existing data**;
- **Rely on multiple measures**;
- **Include assessment results from all grades (3-8, 10) and subjects tested statewide (reading, writing, mathematics, science)**;
- **Incorporate concepts of the federal No Child Left Behind (NCLB) Act and its Adequate Yearly Progress (AYP) system when appropriate**;
- Encourage the continuous improvement of student learning and cooperation among educators;
- Be fair, reasonable, and consistent;
- Be valid and accurate;
- Focus at both the school and district levels;
- Apply to as many schools and districts as possible;
- Use familiar concepts when possible;
- Rely mainly on criterion-referenced measures instead of norm-referenced measures; and
- Provide multiple ways to demonstrate success and earn recognition.

Three assumptions were made during the development of the index.

- Priority schools and districts should be those that are the most challenged in the state – they should meet a "common sense" test as those needing the most support.
- Priority schools and districts would be eligible to receive additional resources to make dramatic improvement in student outcomes through an initiative such as that being developed by Mass Insight. Criteria to be met to receive this support will be specified by the State Board of Education.
- Priority schools and districts would be <u>required</u> to participate in a state-supported initiative, as described by the system being designed by Mass Insight, if offers of additional support are not accepted and substantial improvement did not occur after two years.

## ACCOUNTABILITY INDEX

The proposed index is based on how schools and districts perform on a set of indicators and outcomes. The recommended system uses a matrix of five outcomes and four indicators. The five outcomes are: the results of state assessments in four subjects (reading, writing, mathematics, science) and the "extended" graduation rate (for high schools and districts). These five outcomes are examined using four indicators: (1) achievement, (2) achievement compared to peer schools (this controls for four student characteristics—special education, ELL, low income, and mobility), (3) improvement, and (4) achievement of students from low-income families. This results in 20 different measures, forming a matrix noted in Table 1.

**Table 1: Accountability Matrix**

| | **OUTCOMES** | | | | |
|---|---|---|---|---|---|
| **INDICATORS** | Reading | Writing | Math | Science | Grad Rate |
| Achievement | | | | | |
| Ach. vs. peers | | | | | |
| Improvement | | | | | |
| Ach. of low-inc. | | | | | |

Each cell of the matrix is rated on a 5-point scale (0-4) using a set of fixed benchmarks. These benchmarks reflect the performance in each cell, with 4 being the best outcome. Each of the four subjects is rated using the same set of benchmarks across the entire school (i.e., all subjects have the same set of benchmarks and the assessment results are the aggregate totals for all the tested grades). **The index is the simple average of all 20 ratings.** The higher the index, the better the level of performance of the school or district.

Table 2 shows the four indicators, the five outcomes, and the benchmarks that produce the various ratings. The index ranges from 0.0 to 4.0 and is a number similar to a GPA where 4.0 is the highest score. This numbering scheme also reflects the same system used to describe the levels of performance on the WASL (Levels 0-4). The Learning Index is used to measure the assessment outcome for two indicators: *achievement compared to peer schools* and *improvement* . This index (not to be confused with the accountability index) takes into consideration the percentage of students performing at the five different WASL levels, not just those meeting standard. The Learning Index ranges from 0 to 4, with 4.00 the highest score (similar to a grade point average). This index is explained in detail in Appendix A.

The proposed system does not include AYP results generated for NCLB. AYP results were included in a previous proposal, but feedback from the advisors, members of the Board, and other stakeholders showed a lack of confidence in the validity of AYP results for accountability purposes. The proposed system is more inclusive than the federal system because it includes both writing and science, uses a smaller minimum number for reporting (10 students across the entire school/district), and includes the results of all students, regardless of how long they have been attending school. Nevertheless, various stakeholders believe AYP results still have a role in the state accountability system because (1) the law will likely remain in effect for several more years and AYP results must be calculated, (2) the disaggregation of results by subgroups provides additional details that provide deeper insights into the level of student learning taking place in schools and districts and at individual grade levels, and (3) federal law requires a single accountability system, which

means AYP results need to be included in some way. As a result, the proposed system uses AYP results as one source of data to identify Priority schools and districts once initial index numbers are computed.

**Table 2: Measures and Rating System for Multiple Indicators and Outcomes**

| Indicator | How Outcomes Are Measured | Benchmarks and Ratings | |
|---|---|---|---|
| | | Assessments[1] | Graduation[2] |
| Achievement | • *Assessments:* Percentage of "all" students meeting standard on the WASL/WAAS for all grades assessed<br>• *Graduation:* Extended graduation rate for "all" students | 86-100% .....4<br>70-85.9% .....3<br>55-69.9% .....2<br>40-54.9% .....1<br>< 40% .....0 | ≥ 95% ..... 4<br>85-94.9% ..... 3<br>75-84.9% ..... 2<br>65-74.9% ..... 1<br>< 65% ..... 0 |
| Achievement compared to peers[3] | • *Assessments:* Learning Index of "all" students adjusted for student characteristics (percent of low-income, ELL, special education, and mobile students[4]) for all grades assessed<br>• *Graduation:* Extended graduation rate adjusted for student characteristics (percent low-income, ELL, special education, and mobile students[4]) | > .20 .....4<br>.10 to .20.....3<br>-.099 to .099 ....2<br>-.20 to -.10 .....1<br>< -.20 .....0 | > 12 ..... 4<br>5.01 to 12 ..... 3<br>-5 to 5..... 2<br>-5.01 to -12 .... 1<br>< -12 ..... 0 |
| Improvement | • *Assessments:* Change in the Learning Index from the previous year using results for all grades assessed<br>• *Graduation:* Percentage point change in the extended graduation rate from the previous year | > .12 .....4<br>.051 to .12 .....3<br>-.05 to .05 .....2<br>-.051 to -.12.....1<br>< -.12 .....0 | > 6 ..... 4<br>3.01 to 6 ..... 3<br>-3 to 3 ..... 2<br>-6 to -3.01 ..... 1<br>< -6 ..... 0 |
| Achievement of low-income students | • *Assessments:* Percentage of low-income students meeting standard on the WASL/WAAS for all grades assessed<br>• *Graduation:* Extended graduation rate for all low-income students | 86-100% .....4<br>70-85.9% .....3<br>55-69.9% .....2<br>40-54.9% .....1<br>< 40% .....0 | ≥ 95% ..... 4<br>85-94.9% ..... 3<br>75-84.9% ..... 2<br>65-74.9% ..... 1<br>< 65% ..... 0 |

[1] The same assessment ratings are used for all subjects in all grades.
[2] This outcome only applies to schools that are authorized to graduate students.
[3] This indicator adjusts the outcomes using statistical methods (multiple regression) to control for student characteristics beyond a school's control. Scores are the difference between the school's actual level and the average of the school's peers. Scores above 0 are "beating the odds" and negative scores are below the predicted level. Separate analyses are conducted for the four assessments in elementary, middle, and high schools.
[4] Mobility is the percentage of all students that are not continuously enrolled from October 1 through the testing period, as defined by in OSPI's Core Student Record System.

Tier assignments are determined based on the index score. Schools and districts would initially fall into four tiers based on their accountability index score, with an in-depth analysis of the data and conditions of those in the lowest tier to see if they merit being placed in a fifth (Priority) tier and be eligible to receive more intensive support. A 5-tier system provides sufficient differentiation among schools and districts and corresponds with the numbering of the index system.

Table 3 shows the suggested ranges for the 5-tier system. A suggested descriptive name is given for each tier rather than a numeric designation to avoid confusion about what tier numbers mean. The rating and tier information could be available in a "report card" available to the public, with a set of "stars" indicating the rating so the overall results can be seen at a glance. This intuitive rating symbolism is used in other settings (e.g., rating movies, restaurants, athletes, tourist attractions) and does not require much interpretation. The table also shows the distribution of schools using the criteria shown in Table 2 and data from 2007. A total of 2,046 schools had an index score. Figure 1 shows the index distribution for the 2,046 schools in the analysis. There was little difference in the distribution of schools based on their grades served (i.e., elementary, middle, high).

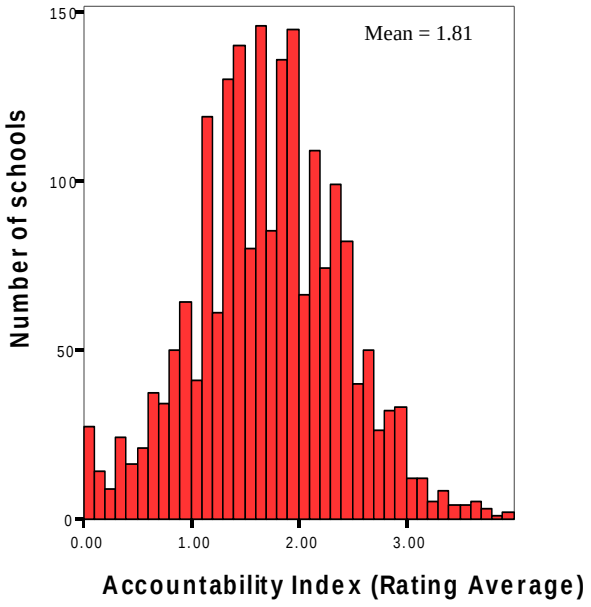
**Table 3: Tier Ranges and 2007 Results** (N=2,046)

| Tier/Suggested Name | Index Range | Number of Schools | Percent of Schools |
|---|---|---|---|
| Exemplary | 3.00 – 4.00 | 72 | 3.5% |
| Good | 2.00 – 2.99 | 664 | 32.5% |
| Adequate | 1.00 – 1.99 | 1,043 | 51.0% |
| Struggling | 0.00 – 0.99 | 267[2] | 13.0% |
| Priority (eligible for Innovation Zone)[1] | 0.00 – 0.99 | TBD | TBD |

[1]Schools and districts in the lowest tier would be determined after an in-depth analysis of quantitative and qualitative information.

[2]Of these schools, 103 (39% of this group) were alternative schools or served other special populations. There were about 83,000 students enrolled in the schools in this tier in 2007 (about 8.3% of all students statewide). About 70,000 students attended "regular" schools that were in this tier.

**Figure 1: Distribution of Schools by Index Score**[**]




[**] All the schools with an index of 0.00 served special populations (correctional facilities, alternative schools, dropout recovery programs), and most had fewer than 10 assessed students so their results would not be reported. The lowest index for a regular school was 0.13, but this school made substantial gains in 2008.

Tables 4 and 5 give examples of how the individual ratings generate the index/tier assignment for two actual schools using results available from 2007. The schools' final index is shown graphically relative to the entire continuum. The tiers and average ratings are noted in colors that correspond to the colors used for the WASL levels on the OSPI Report Card. The results could be made public as part of the OSPI Report Card (the format of the presentation must still be determined). Results presented in this type of "dashboard" give policymakers, educators, and the public a quick snapshot of where a school is strong and weak, its overall rating, and where it falls within the tier. It also provides transparency about how the index number is determined.

- The *high school* described in Table 4 is located in a medium-sized suburb of a large city with fewer low-income students than the typical high school in the state. Its WASL scores had been about the state average in most subjects but both reading and math scores dropped dramatically from 2006 levels. Like many high schools, it has low math and science scores. It also has lower scores than high schools serving similar students, and the performance of its low-income students was below that of "all" students in four subjects. Its graduation rate is fairly high, even when compared to its peers, the rate improved substantially from the previous year, and surprisingly, low-income students had a higher rate than the "all" students rate. Its index of 1.65 puts it close to the middle of the "adequate" tier, which is probably worse than educators and community members expected.
- The *elementary school* described in Table 5 is located in a medium-sized city with above-average levels of low-income, ELL, and mobile students. Its WASL scores are well above the state average in several grades but below the state average in one grade. It had sharp declines from very high WASL scores the previous year, resulting in low improvement ratings in 3 subjects. Its reading and writing scores are still quite high and its scores are very high compared to schools serving similar students. Low-income students had the same rating as "all" students in three subjects but were lower in writing. The graduation rate does not apply. Its index of 2.13 is slightly above the middle of the index scale and in the lower end of the "good" tier.

**Table 4: "Actual" High School, 2007**

| Indicator | Reading | Writing | Math | Science | Grad Rate | Average |
|---|---|---|---|---|---|---|
| Achievement | 3 | 3 | 1 | 0 | 3 | 2.00 |
| Ach. vs. peers | 1 | 1 | 1 | 1 | 3 | 1.40 |
| Improvement | 0 | 2 | 0 | 2 | 4 | 1.60 |
| Low-inc. ach. | 2 | 2 | 0 | 0 | 4 | 1.60 |
| Average | 1.50 | 2.00 | 0.50 | 0.75 | 3.50 | **1.65** |
| Achievement | *** | *** | * | | *** | |
| Ach. vs. peers | * | * | * | | *** | |
| Improvement | | * | | ** | **** | |
| Low-inc. ach. | ** | ** | | | **** | |

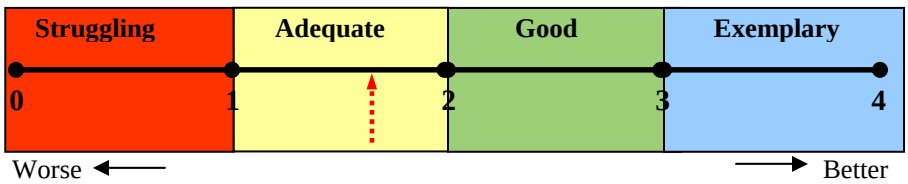
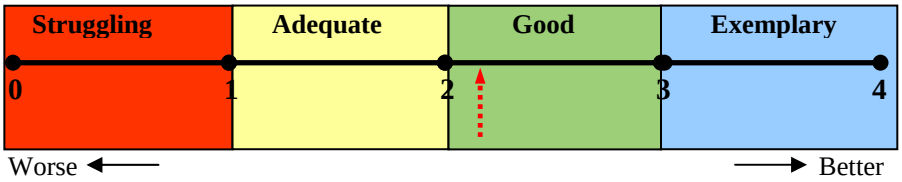


**Table 5: "Actual" Elementary School, 2007**

| Indicator | Reading | Writing | Math | Science | Grad Rate | Average |
|---|---|---|---|---|---|---|
| Achievement | 3 | 3 | 2 | 0 | | 2.00 |
| Ach. vs. peers | 4 | 4 | 4 | 3 | | 3.75 |
| Improvement | 0 | 2 | 1 | 1 | | 1.00 |
| Low-inc. ach. | 3 | 2 | 2 | 0 | | 1.75 |
| Average | 2.50 | 2.75 | 2.25 | 1.00 | | **2.13** |
| Achievement | *** | *** | ** | | | |
| Ach. vs. peers | **** | **** | **** | ** | | |
| Improvement | | * | ** | * | | |
| Low-inc. ach. | *** | ** | ** | | | |




## DISTRICT ACCOUNTABILITY

The proposed system would hold districts accountable using the same rules, indicators, and outcomes that are used for school accountability. The results would be based on districtwide data for all grades rather than being disaggregated by grade bands (elementary, middle, high). District results are more likely to be made public when using the combined results for all grades—only five extremely small districts, with a combined total of 34 students, had fewer than 10 students in their tested grades in 2007. Financial data, which is available only at the district level on a consistent basis, is used as an additional independent variable in the district-level regression to control for the amount of total operating revenue per pupil available to the district. The same type of deeper analyses would occur for districts that have an index number in the lowest tier in order to determine if they merit receiving extra support, just like the process used for schools.[1] This closer look would also include examining the percentage of schools and number of students that are found in the lowest tier and the consistency of problems in a particular set of grade bands or subjects. Since more information is available at the district level, district accountability could include additional measures besides the 20 in the matrix. Moreover, other data could be used when analyzing districts and their peers, such as unemployment rates, crime rates, per capita income, and tax base if this information is available at the district level.

Various tables and charts can illustrate the district results. Table 6 and Figure 2 show how all the results for a district can be shown across multiple years to show trends over time. (State results are used, and the data in shaded cells of the table are not available.) Figure 3 shows the distribution of the *number of schools* by tier for an actual district. Figure 4 shows the *percentage of students* enrolled at those schools. (One alternative high school has relatively few students.)

[1] Districts are not required to have an improvement plan unless they are in district improvement. The State Board could require districts to have such a plan, just like schools. A review of the district plan (its quality and use) could be part of the more intensive analysis of district conditions.

**Table 6: Showing Results Over Time (All Grades)**

| | *YEAR* | | | | |
|---|---|---|---|---|---|
| *Indicator/Outcome* | ***2003*** | ***2004*** | ***2005*** | ***2006*** | ***2007*** |
| **Achievement** | **1.25** | **1.25** | **1.60** | **1.60** | **2.00** |
| Reading | ** | ** | *** | *** | *** |
| Writing | ** | ** | ** | ** | *** |
| Math | * | * | * | * | ** |
| Science | | | | | |
| Grad. rate | | | ** | ** | ** |
| **Ach. vs. peers**[1] | **2.00** | **2.00** | **2.00** | **2.00** | **2.00** |
| Reading | ** | ** | ** | ** | ** |
| Writing | ** | ** | ** | ** | ** |
| Math | ** | ** | ** | ** | ** |
| Science | ** | ** | ** | ** | ** |
| Grad. rate | | | ** | ** | ** |
| **Improvement** | **2.50** | **3.67** | **3.25** | **2.60** | **1.80** |
| Reading | ** | **** | **** | ** | ** |
| Writing | | | *** | **** | ** |
| Math | *** | **** | *** | ** | ** |
| Science | | *** | *** | ** | ** |
| Grad. rate | | | | *** | * |
| **Low-income Ach.**[2] | | **0.50** | **0.80** | **1.00** | **1.20** |
| Reading | | * | ** | ** | ** |
| Writing | | * | * | ** | ** |
| Math | | | | | * |
| Science | | | | | |
| Grad. rate | | | * | * | * |
| **Account. Index** | | **1.73** | **1.84** | **1.80** | **1.75** |

[1]This indicator does not apply in this example because the state has no peer, so a middle rating is given in each year for all outcomes.
[2]Student counts for subgroups are not available for 2003, so no rating was determined and no index is calculated.

**Figure 2: Average Ratings, 2003-2007**

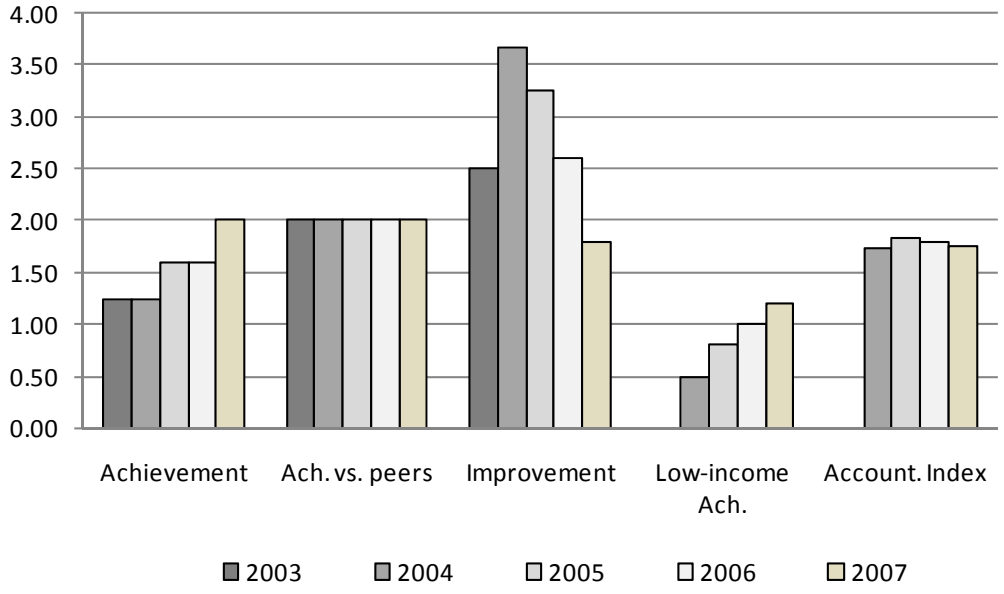


**Figure 3: Distribution of *Schools* by Grade Level and Tier in "Actual" District**

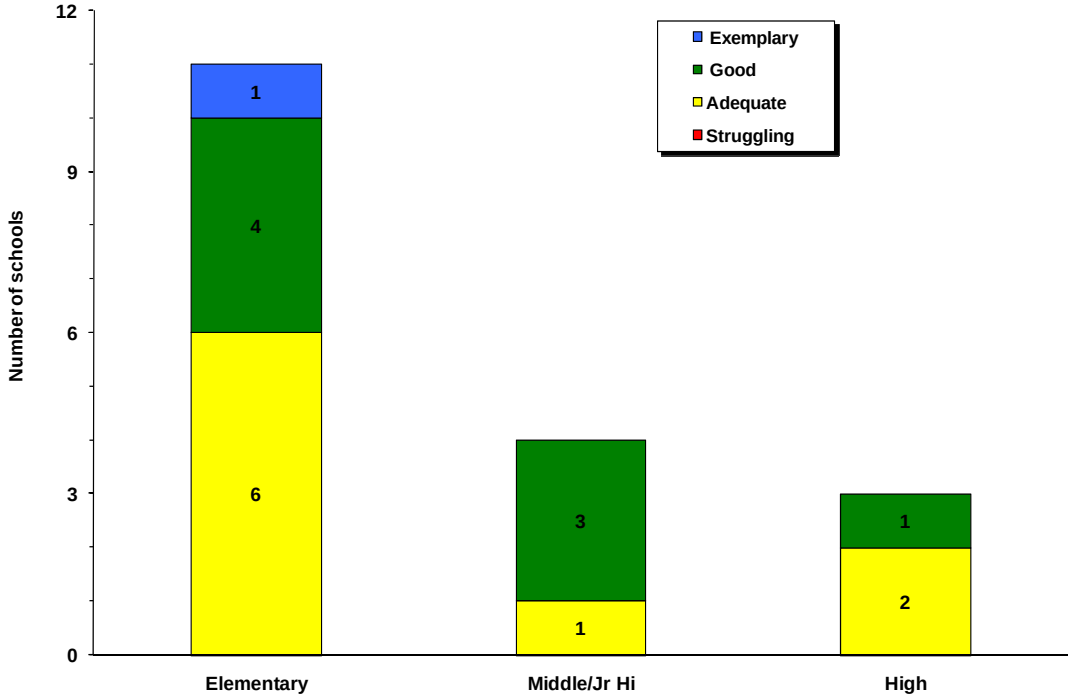



**Figure 4: Distribution of *Students* by Schools in Tiers and Grade Level in "Actual" District**

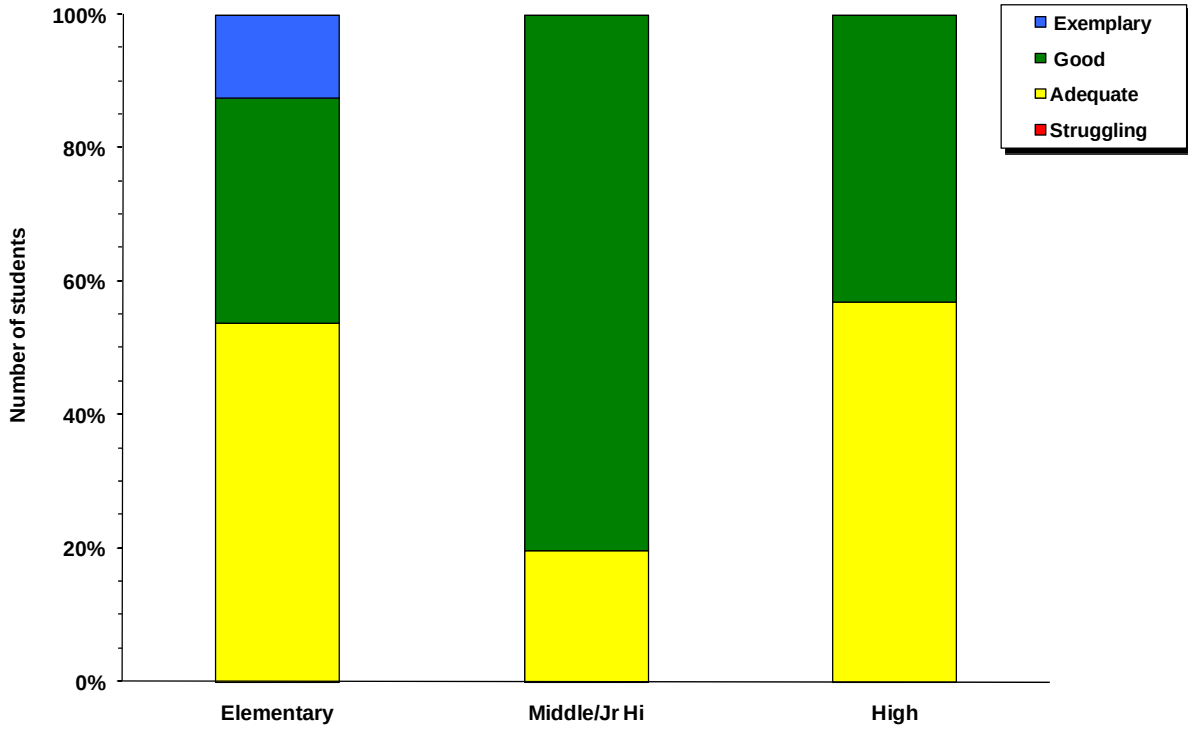
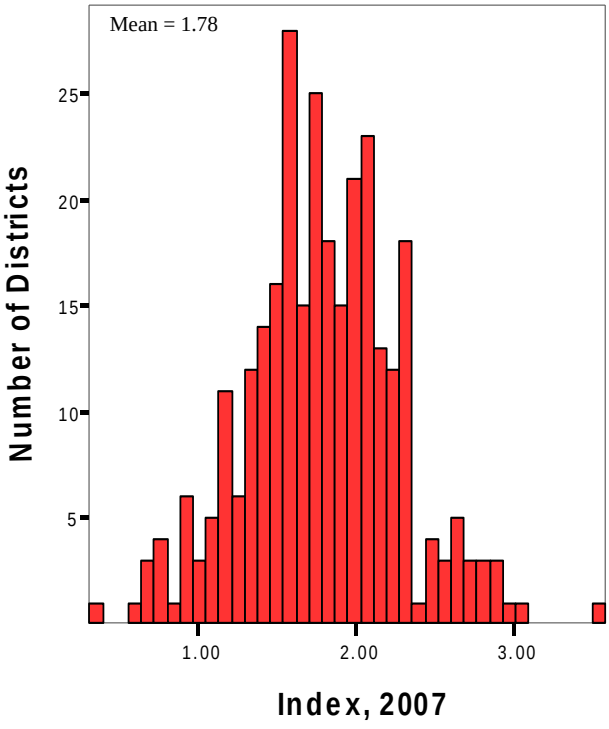


Table 7 shows the district results using the same criteria and rating system used for schools. Districts are more tightly clustered in the distribution than schools, with fewer districts in the top and bottom tiers (see Figure 5).[2] Figure 6 provides a different view of the district index results. Of the 16 districts in the "struggling" tier, the average size was about 1,000 students (the median was slightly more than 400 students). Half of the 16 districts made AYP in part because the AYP targets were relatively low in 2007, the margin of error is large for small districts, and many of the student groups in the smaller districts had fewer students than the required minimum to make a AYP determination.

**Table 7: Tier Ranges and 2007 Results for Districts** (N=296)

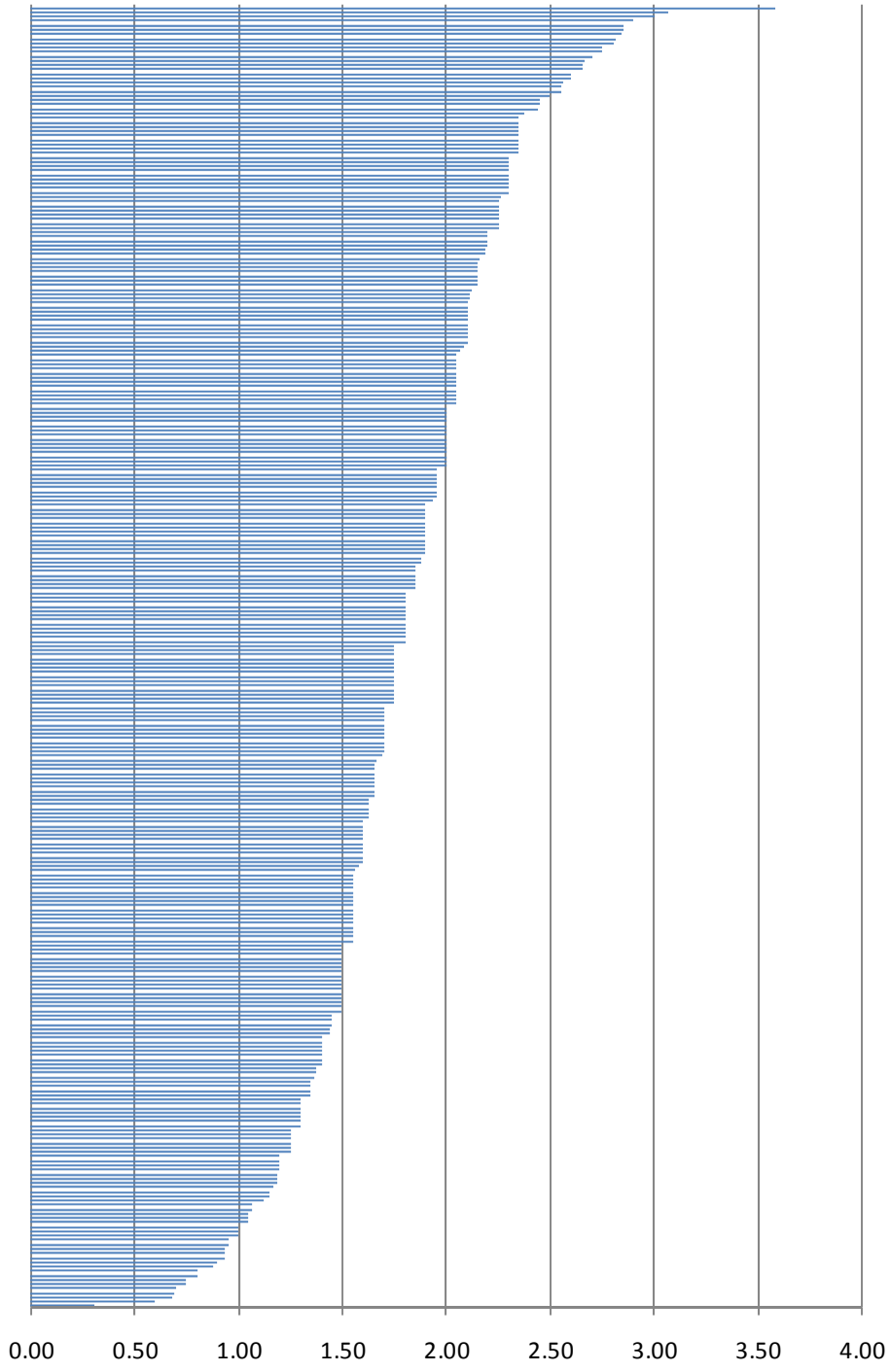
| Tier/Suggested Name | Index Range | Number of Districts | Percent of Districts |
|---|---|---|---|
| Exemplary | 3.00 – 4.00 | 3 | 1.0% |
| Good | 2.00 – 2.99 | 102 | 34.5% |
| Adequate | 1.00 – 1.99 | 175 | 59.1% |
| Struggling | 0.00 – 0.99 | 16 | 5.4% |
| Priority (eligible for Innovation Zone) | 0.00 – 0.99 | TBD | TBD |

**Figure 5: Distribution of Districts by Index Score**




[2] District results do not include several types of schools. For example, correctional institutions, tribal schools, contract schools, and schools serving more than 50% of students outside the district boundary. The aggregation rules using in these calculations are the same as those used by OSPI when calculating district results.

**Figure 6: Distribution of All Districts by Index Score**



## IDENTIFYING PRIORITY SCHOOLS AND DISTRICTS (LOWEST TIER)

The process for identifying Priority schools and districts would begin when OSPI computes the index in mid-August using the most recent data and prepares a set of preliminary results. Given the relatively large number of schools that may fall into the "struggling" tier,[3] the schools must be screened to eliminate those that clearly should not fall into the Priority tier, which would reduce the number of schools and districts that require a deeper analysis. When OSPI and SBE staff are confident the index has been calculated correctly, OSPI staff review the index results for each school and district that falls in the "struggling" tier, and then sort them into two categories:

(1) Schools/districts that are *Not for Priority designation* are those that have not been in the "struggling" tier in the past two years or have obvious data problems that affected their results (e.g., errors in reporting the number of graduates, missing data for ELL, special education, and low income students that can affect the results of the "peers").

(2) The remaining schools/districts are placed in a *Possible Priority tier* category pending a deeper analysis.

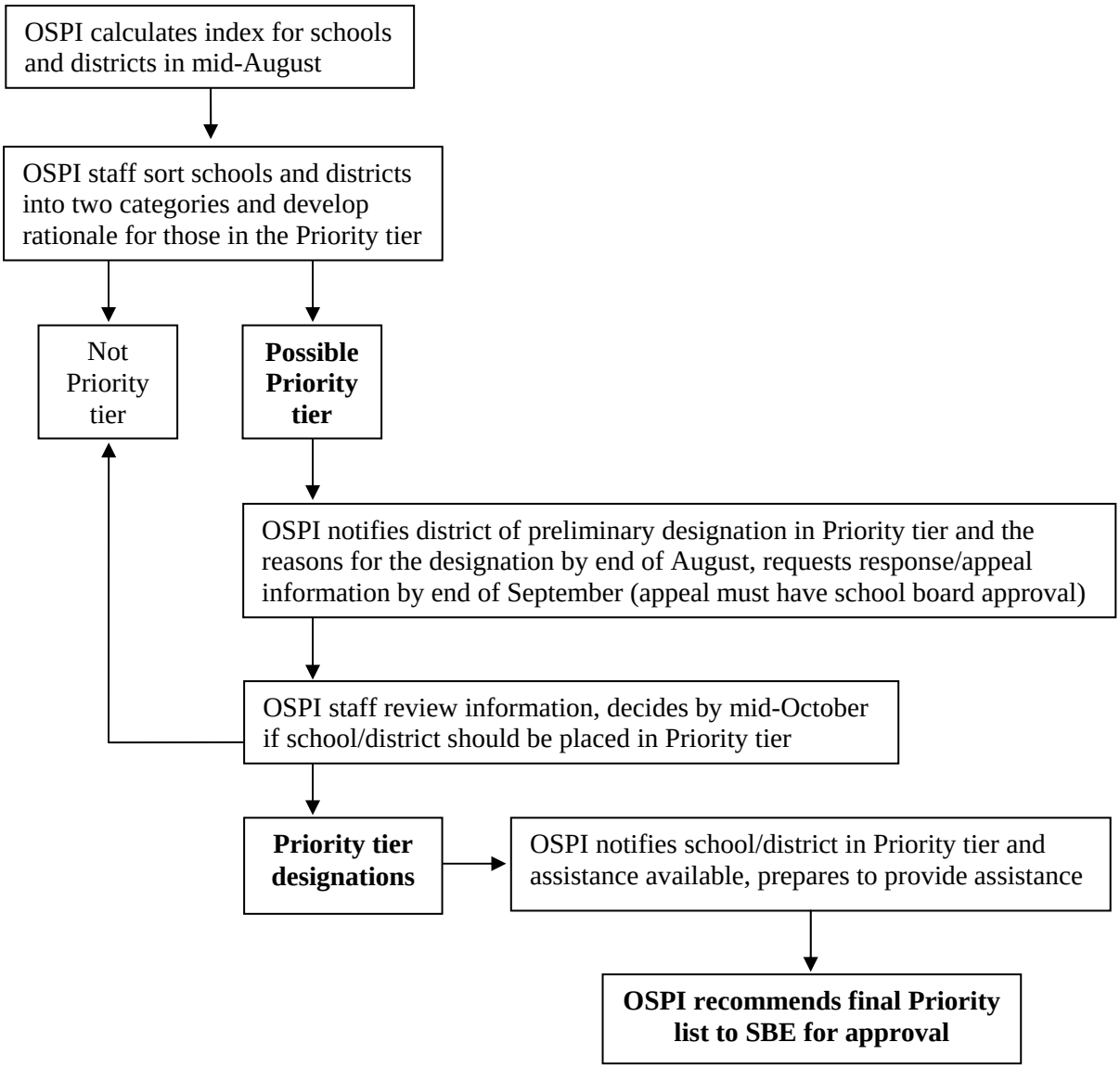
OSPI staff will conduct a deeper analysis available data for the schools and districts placed in the possible Priority tier category. This may require contacting the district and/or local ESD to get more information. A comprehensive list of quantitative and qualitative data was developed that could be used to help determine which schools in the "struggling" tier should fall into the "Priority schools" tier (see Appendix B). Given the comprehensive nature of the list and the limited capacity to analyze all the data for every school and district in the "struggling" tier, the list was refined to determine which were the most important data to analyze. The data that would be initially reviewed at this exploratory phase fall into four general areas:

- **Contextual Data:**
  - Type of school
  - Changes in student population
  - Programs served by the school
  - Level of student mobility
- **WASL/WAAS Results**
  - Trends over multiple years for each subject area
  - Subgroup trends
  - Results for students who have been enrolled for at least two years
- **AYP Results:**
  - Distance from the annual goal
  - Type of cells not making AYP
  - Percentage of cells not making AYP
- **Other Data:**
  - Graduation and dropout rates for subgroups
  - Student/teacher ratio
  - Teacher education and experience levels
  - Funding from local levies/bonds and outside sources
  - Recent changes in leadership (key central office staff and principals) and teachers

[3] The number will still be far fewer than those not making AYP or identified for "improvement" under NCLB.

Based on this review, the schools and districts are sorted again into the same two categories—not for Priority designation and possible Priority tier. By the end of August, districts of schools placed in the possible Priority tier are notified of the possible designation and given the reasons why designation is possible. If required by federal law, this initial list would be made public. During the month of September, the district/school is given a chance to avoid the Priority designation by providing more information that would explain the low index results, and it could provide more favorable results (e.g., feeder school information, results of district assessments, personnel changes, type of interventions made to date) and any plans being made for the future. Any appeal needs to have school board approval. OSPI reviews the additional information, and by mid-October, it determines the schools and districts placed in the Priority tier. Figure 7 provides a flow chart of this process.

**Figure 7: Process for Identifying Priority Schools and Districts**



## INTEGRATING THE SYSTEMS

Federal law requires states to have a single accountability system. Many states combine their state accountability system with the federal system described by NCLB. The details for integrating the federal and state system must still be determined. Washington state can pursue two options to meet this requirement.

1. The preferred approach is to request that the proposed system be used in place of the current system. A new administration may provide more flexibility to states to design alternative accountability systems and approve them if they meet certain requirements. The proposed system has many desirable features that could make it a viable alternative to the current rules used to measure AYP.

2. If Washington is not allowed to use the proposed system to replace the current AYP system, the results of the calculations from the two systems could still be used when determining the type and level of assistance the states provide. Those that fall into "improvement" status under AYP would still face the required sanctions. However, schools that do not make AYP and fall into school improvement may also achieve relatively favorable index results. In these cases, the amount of assistance the state provides would be minimal. On the other hand, some schools will make AYP and not be in school improvement, but they may have relatively low index results. In these cases, state funds could be used to focus assistance in areas of greatest need. Regardless of the results from the two systems, the state must be sure to clarify what happens when schools and districts fall into the various AYP categories and state tiers and make every effort to minimize confusion that could occur about the two ways for measuring accountability. Appendix C provides an overview of the current assistance system being used by OSPI to help schools and districts that are in "improvement" status.

## RECOGNITION

Three of the guiding principles for developing the accountability system are to (1) provide multiple ways to demonstrate success and earn recognition, (2) rely mainly on criterion-referenced measures, and (3) simple to understand. The proposed recognition system is consistent with these principles. It will use the results from the accountability matrix and provide recognition in each of the 30 cells of the matrix: each of the 20 "inner" cells of the matrix, the average of the four indicators and five outcomes, and the overall index. A minimum rating of 3.00 is required to receive recognition in the 20 "inner" cells, and a minimum of 2.75 rating is needed to receive recognition for the "averaged" cells (see Table 8). Any cell with a 3.5 or above would receive recognition "with honors." The ratings will be calculated every year, and recognition is given when the two-year average rating meets the minimum requirement. This ensures that recognition is given for sustained exemplary performance.

**Table 8: Minimum Requirements for Recognition**

| Indicator | Reading | Writing | Math | Science | Grad Rate | **Average** |
|---|---|---|---|---|---|---|
| Achievement | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.75 |
| Ach. vs. peers | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.75 |
| Improvement | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.75 |
| Low-inc. ach. | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.75 |
| **Average** | 2.75 | 2.75 | 2.75 | 2.75 | 2.75 | 2.75 |

Table 9 and Figure 8 show how many of the 2,046 schools would have received awards if the proposed system was in place in 2007 (district results were not calculated). The largest number of schools would have received recognition in just one or two of the 30 areas, and 16% would not have received any recognition. At the other extreme, about 14% of schools would have received recognition in 10 or more areas, and 2 schools would have received recognition in 22 of the 30 cells of the matrix.

**Table 9: Number of Schools of Distinction, by Number of Recognitions (2007)**

| Number of recognitions at a school | Number of schools | Pct of all schools | Cumulative percent |
|---|---|---|---|
| 0 | 330 | 16.1% | 16.1% |
| 1 | 338 | 16.5% | 32.6% |
| 2 | 260 | 12.7% | 45.4% |
| 3 | 185 | 9.0% | 54.4% |
| 4 | 169 | 8.3% | 62.7% |
| 5 | 143 | 7.0% | 69.6% |
| 6 | 104 | 5.1% | 74.7% |
| 7 | 85 | 4.2% | 78.9% |
| 8 | 77 | 3.8% | 82.6% |
| 9 | 64 | 3.1% | 85.8% |
| 10 | 59 | 2.9% | 88.7% |
| 11 | 55 | 2.7% | 91.3% |
| 12 | 33 | 1.6% | 93.0% |
| 13 | 41 | 2.0% | 95.0% |
| 14 | 18 | 0.9% | 95.8% |
| 15 | 20 | 1.0% | 96.8% |
| 16 | 14 | 0.7% | 97.5% |
| 17 | 18 | 0.9% | 98.4% |
| 18 | 12 | 0.6% | 99.0% |
| 19 | 10 | 0.5% | 99.5% |
| 20 | 6 | 0.3% | 99.8% |
| 21 | 3 | 0.1% | 99.9% |
| 22 | 2 | 0.1% | 100.0% |

**Figure 8: Number of Schools of Distinction, by Number of Recognitions (2007)**

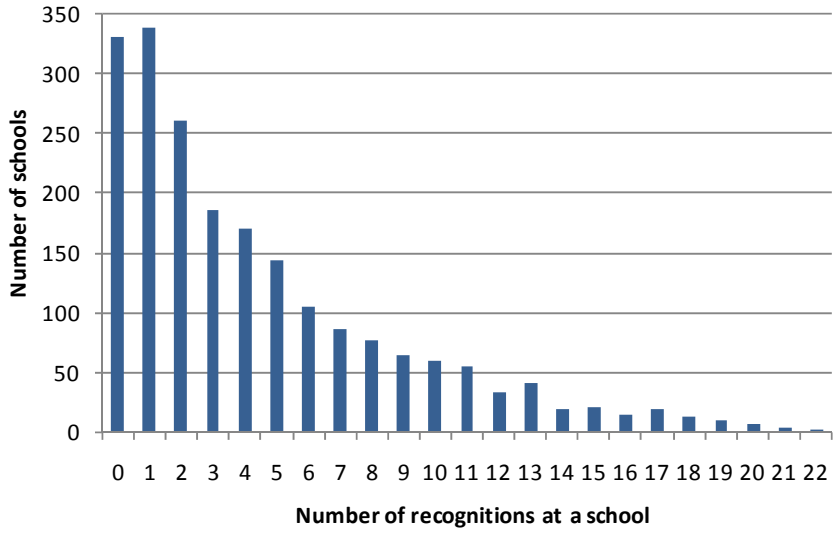


Table 10 shows the number of schools that met the recognition criteria in each area in 2007. The largest number of schools (52%) met the criteria for reading achievement. Achievement in math, science, and among low-income students had fewer schools meeting the criteria. Only 4% had an overall average of 2.75 on the accountability index over the 2-year period. Although schools would have received recognition in a total of 9,082 areas, this represents less than 15% of the total maximum possible (30 cells x 2,046 schools). Figure 9 shows the percentage of all schools that would have received each type of recognition.

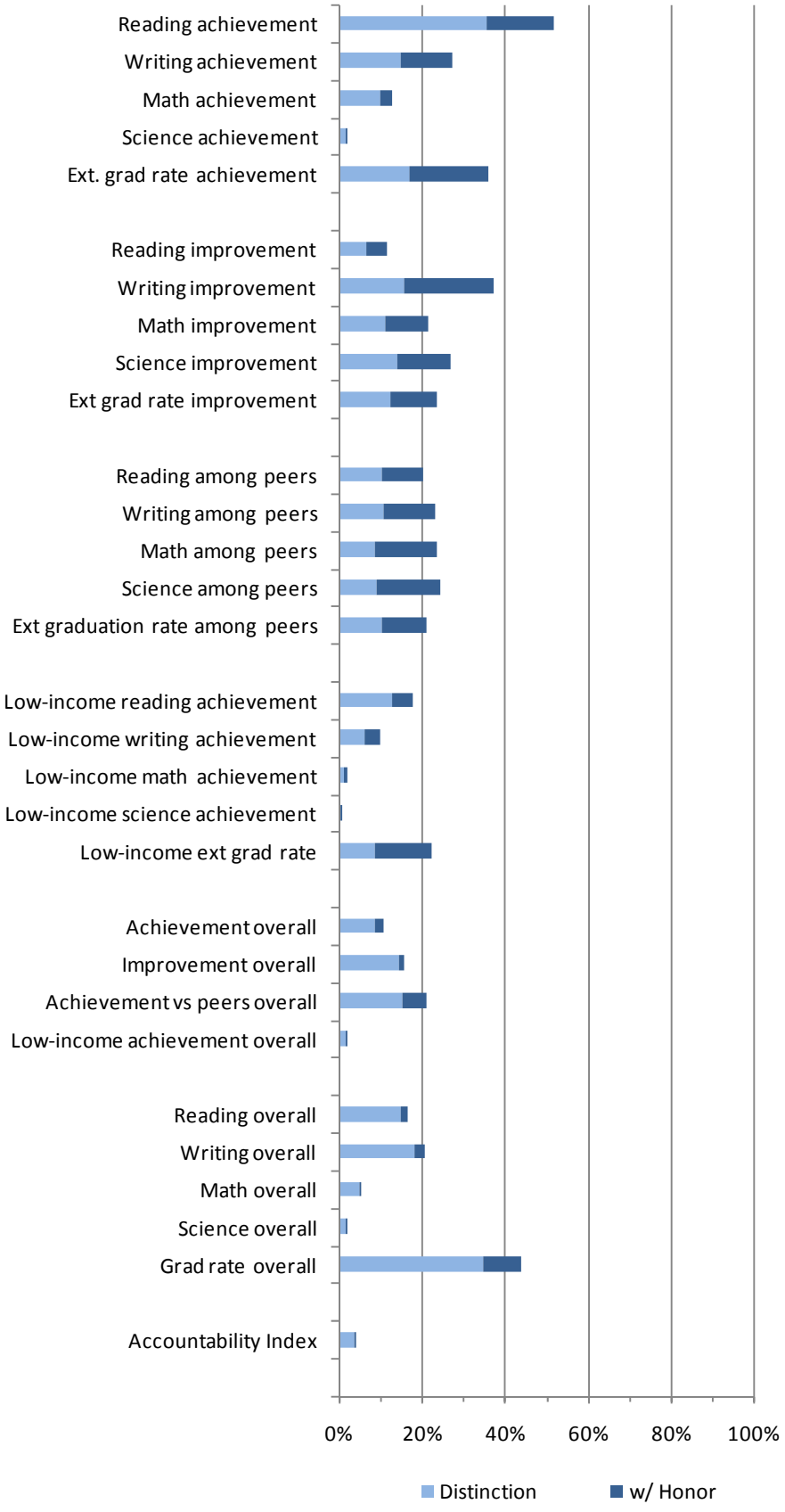
**Table 10: Distribution of Schools of Distinction, by Type of Recognition (2007)**

| Type of Recognition | # of "Schools of Distinction" | # of "Schools of Distinction" *with Honor* | Total # of schools recognized | Pct of all schools** |
|---|---|---|---|---|
| Reading achievement | 727 | 330 | 1,057 | 51.7% |
| Writing achievement | 309 | 255 | 564 | 27.6% |
| Math achievement | 204 | 60 | 264 | 12.9% |
| Science achievement | 37 | 9 | 46 | 2.2% |
| Ext. grad rate achievement | 75 | 83 | 158 | 36.0% |
| Subtotal, Achievement[1] | 1,352 | 737 | 2,089 | |
| Reading improvement | 135 | 100 | 235 | 11.5% |
| Writing improvement | 322 | 446 | 768 | 37.5% |
| Math improvement | 230 | 209 | 439 | 21.5% |
| Science improvement | 286 | 265 | 551 | 26.9% |
| Ext grad rate improvement | 54 | 50 | 104 | 23.7% |
| Subtotal, Improvement[1] | 1,027 | 1,070 | 2,097 | |
| Reading among peers | 210 | 210 | 420 | 20.5% |
| Writing among peers | 221 | 254 | 475 | 23.2% |
| Math among peers | 176 | 312 | 488 | 23.9% |
| Science among peers | 191 | 313 | 504 | 24.6% |
| Ext graduation rate among peers | 46 | 46 | 92 | 21.0% |
| Subtotal, Peers[1] | 844 | 1,135 | 1,979 | |
| Low-income reading achievement | 259 | 105 | 364 | 17.8% |
| Low-income writing achievement | 128 | 78 | 206 | 10.1% |
| Low-income math achievement | 26 | 17 | 43 | 2.1% |
| Low-income science achievement | 5 | 4 | 9 | 0.4% |
| Low-income ext grad rate | 38 | 61 | 99 | 22.6% |
| Subtotal, Low Income[1] | 456 | 265 | 721 | |
| Achievement overall | 179 | 41 | 220 | 10.8% |
| Improvement overall | 297 | 29 | 326 | 15.9% |
| Achievement vs peers overall | 311 | 125 | 436 | 21.3% |
| Low-income achievement overall | 30 | 7 | 37 | 1.8% |
| Reading overall | 306 | 30 | 336 | 16.4% |
| Writing overall | 374 | 48 | 422 | 20.6% |
| Math overall | 103 | 8 | 111 | 5.4% |
| Science overall | 33 | 6 | 39 | 1.9% |
| Grad rate overall | 153 | 40 | 193 | 44.0% |
| Accountability Index | 75 | 1 | 76 | 3.7% |
| Total[1] | 5,540 | 3,542 | 9,082 | |

** N=2046 for academic measures; N=439 for extended graduation rate measures

[1] Duplicated count

**Figure 9: Percentage of Schools of Distinction, by Number of Recognitions (2007)**



This system of recognition would supplement and could replace some types of recognition currently in place. The federal government provides funding for three types of awards, primarily for schools receiving Title I funds. OSPI also provides awards but no funding as part of the recognition.[4] Schools and districts that receive recognition in the proposed system would not be compensated monetarily, although exceptions could be made. In its compensation proposal to the Basic Education Finance Task Force, OSPI has recommended that schoolwide financial rewards be given each year when a school reaches a certain level of improvement. The proposed recognition system could be used as a basis for these rewards. For example, schools that have an average of at least 3.0 for overall improvement could be given a schoolwide financial bonus. In 2007, about 8% of the *schools* statewide would have qualified for this bonus (15% of the *districts* averaged 3.0 or better in the improvement cells). The amount of the bonus suggested by OSPI was $20 to $50 per student FTE. Other types of recognition, with or without financial awards, could be developed. These could be available to all that meet certain criteria and/or be competitive in nature.

* * * * * *

The proposed accountability system needs to be flexible. Changes in NCLB requirements, graduation requirements, the graduation rate formulas, the assessment system, and content standards may have an impact on some measures, which may require changes to the system. And as data systems improve statewide and more information becomes available, other indicators could be added to the system and other more sophisticated analyses could be used (e.g., growth models). These changes could be in the form of additional columns in the matrix (e.g., college eligible rates) or additional factors outside the matrix that could be included when calculating the index (e.g., funding amount of local levies).

Appendix A provides more details about how the index is calculated. Appendix B provides a list of possible data that could be used to identify Priority schools. Appendix C gives an overview of the current state assistance system that is funded primarily by the federal government. Appendix D lists the names of those who provided advise and feedback during the development of this proposal.

---

[4] **Blue Ribbon** schools are nominated by OSPI and selected by the U.S. Department of Education based on high academic performance. In order to be selected, nominated schools must provide detailed information about their school, they can be any type of school (including private schools), and they must make AYP in the year of the nomination and the following year. For the **Academic Achievement Award** program**,** Title I Part A schools that met AYP for three consecutive years in math and/or reading can apply for recognition of improving student achievement in one or both content areas. Up to nine schools can receive an award of $10,000. The application provides details about successful math and/or reading strategies, and these strategies are showcased at state conferences and on OSPI's website in order to assist other schools. For the **Distinguished Schools Award**, four Title I Part A schools are selected, two in the national category and two in the state category. Schools must apply for this award, which focuses on either exceptional student performance for two or more years or significant progress in closing the achievement gap. National award winners receive $10,000 while state award winners receive $5,000. OSPI began recognizing **Schools of Distinction** in 2007 based on improvement over an extended period of time and achievement that exceeds the state average. Only the top 5% of schools received this award. Finally, OSPI has been giving **Improvement Awards** since 2004 to schools and district that make at least a 10% reduction in the percentage of students not meeting standard in reading, writing, and math in grades 4, 7, and 10. Wall plaques with metal plates for updates are provided to those receiving this award. In 2007, there were 1,255 schools that received a total of 2,190 awards in the three grades and subjects; 241 districts that received a total of 804 awards in the three grades and subjects. OSPI does not provide any recognition or results based on how schools or districts compare to their peers.

# APPENDIX A

## INDICATORS AND OUTCOMES

This appendix provides more detailed information about the proposed accountability index. It also includes information about how the indicators and outcomes were selected and how the index number is calculated.

### SELECTION OF INDICATORS AND OUTCOMES

One of the guiding principles for the accountability system is the use of multiple measures. The advisors (see Appendix D) recommended using four indicators and five outcomes, resulting in a 4x5 matrix with 20 outcomes. Other indicators and outcomes were discussed besides the WASL and graduation rates, and other outcome data were desired in order to have multiple measures. However, no other reliable and accurate data are available statewide that is collected in the same manner.

The index is achieved by using the simple average of the ratings across the 20 outcomes. The graduation rate is not applicable for elementary and middle schools, but these types of schools have multiple grades with WASL results that generate the ratings. By using averages, schools without data for some indicators are still included in the system and a separate system is not needed for different types of schools to generate the index.

The advisors preferred a system that uses fixed criteria rather than norm-referenced measures in order to keep the measures simple and to avoid changing goals over time and the use of measures (e.g., standard deviations) that vary by subject. This means that recognition would be given when schools meet certain criteria, and there would not be a limit to how many schools can be recognized (unlike the Schools of Distinction which only recognized the top 5% based on improvement). With fixed criteria in place, a school and district would know in advance what it needed to do to receive recognition, regardless of how others perform. It would also encourage cooperation among educators because they would not be in competition with one another for recognition.

The advisors discussed other types of analyses that could provide more accurate results (e.g., structural equation modeling, hierarchical linear modeling, value-added growth models). However, these methods were not selected because they lack transparency, are overly complex, and are not calculated easily at the school and district levels due to capacity and software limitations.

All stakeholder groups believed the federal AYP system is not a valid way to identify schools for recognition and additional support. The advisors felt the current system is too complex, has too many adjustments, and is neither transparent nor fair in its accountability determinations. Moreover, AYP is almost entirely punitive in nature and does not include two subjects (writing and science) that are assessed in a standardized manner statewide, which has resulted in a narrowing of the curriculum. AYP's narrow emphasis on students who meet standard has often resulted in more focused help being given to students that perform near that cut point (known as the "bubble kids") and at the expense of students who are farther above and below that level of performance.

The proposed system is preferred because it is more inclusive and less complex than the federal AYP system. The ratings are based on the results for all students, including those who are not "continuously enrolled" since October 1. No margin of error is used, and the minimum N is 10 across the entire school/district (rather than a grade) in order to increase the chance that very small schools and districts (e.g., those with less than 10 students in a grade) are included in the accountability system. For example, a K-6 school that has only 4 students in each tested grade (grades 3-6) would have a total of 16 students with assessment results and would therefore be included in the system. (Grade-level results are not reported when there are fewer than 10 students in a grade in order to keep the results confidential.) Grade configurations are not an issue when calculating the results because the same benchmarks are used for each grade and subject (AYP uses grade bands of 3-5, 6-8, and 10 with separate results generated for each grade band, regardless of the school's grade configuration). The current AYP system for holding districts accountable is even more complex than the school accountability system. It has different rules and sometimes produces results that are confusing and at odds with its school-level results (e.g., a district might not make AYP but all its schools do and vice versa). A district's size is the major determinant in its AYP results—only two districts with fewer than 1,000 students are in improvement status. The proposed district accountability system is essentially the same as the system for schools, which makes it relatively easier to understand and compute.

## USING THE INDEX

The results from the 20 ratings create an index number for each school and district based on the average rating. Schools and districts are assigned to a "tier" based on their index number.

- Those with the highest index numbers, from 3.00 to 4.00, are in the "exemplary" tier.
- Those with an index of 2.00 to 2.99 are in the "good" tier.
- Those with an index of 1.00 to 1.99 are in the "adequate" tier.
- Those with an index below 1.00 are in the "struggling" tier.

Schools should not be compared and judgments should not be made about school quality based solely on their overall index score. Even though the index uses multiple measures, some schools have missing data that can affect their index number. Moreover, schools that administer assessments with lower scores overall (e.g., science and math) will tend to have a lower index score than those that do not. For example, schools serving grades 5, 8, and 10 give the science WASL, and these results tend to be very low compared to the other subjects. So a K-4 school will likely have a higher index score than a K-5 or K-8 school. Schools serving very few students may have more volatile ratings from year to year. As a result, the index is only comparable across schools that serve the same grades. In addition, the index does not reflect how close a school may be to the benchmarks—small differences in results could still generate different ratings (e.g., 85%=3 and 86%=4). The lack of vertical alignment of the assessments presents another complicating factor when making comparisons across schools that serve different grade levels.

Given the different types of schools being rating, school results should be reported for similar types of schools. The six suggested categories for reporting the results are as follows:

- *elementary schools* (those serving from kindergarten up to grade 6)
- *middle/junior high schools* (those serving only 6,7 or 8)
- *high schools* (grades 9 or 10 to 12)
- *comprehensive schools* (e.g., K-8, K-12)

- *schools serving special populations* (alternative schools, correctional facilities, those primarily serving ELL students and those with disabilities, private schools on contract)
- *small schools* (those which have their results suppressed because the y have fewer than 10 assessed students).

Many districts have only one school. As a result, their index, tier, and recognition would be the same. This has implications for how the state structures the consequences of the accountability system (either with assistance or recognition).

The accountability system will need to remain flexible. Changes in NCLB requirements (e.g., number of tested grades), graduation requirements, the method for calculating the graduation rates, the assessment system (e.g., moving to end-of-course exams in math, adjustments to cut scores), and content standards (e.g., science) may have an impact on some measures, which may require adjustments to the accountability system. Moreover, as data systems improve statewide and more information becomes available, other indicators can be added to the system[5] and other more sophisticated analyses could be used (e.g., growth models). Other measures of improvement could be used (computing expected change, percent increases). Changes could also be in the form of additional columns in the matrix (e.g., college eligible rates) or additional factors outside the matrix that could be included when calculating the index or peer results (e.g., funding amount of local levies).

## CALCULATION METHODS

To calculate the achievement measures, student-level data were used and aggregated to the school and districts levels. This provides more accurate results than using aggregated school and district results. Moreover, using student-level data allows for the aggregation of results from the grade level that would be suppressed because the number of students assessed was less than 10. Results are only suppressed when there are fewer than 10 students assessed in the combined grades.[6] Students who took the alternate assessments (WAAS) were included in the calculations, as were students who previously passed (this relates mainly to high school students that met standard while in grade 9, but it also applies to students that are retained). Students who met standard in a previous year did not have their level included in the student-level database, so they were considered to have performed at Level 3. Students who were exempted from taking the assessments (i.e., those with excused absences and medical exemptions, first-year ELL students, home-based and private school students) were not included in the calculations.

When computing the index, all the ratings are counted equally (i.e., they are not weighted). Averages are computed only for cells of the matrix that had data (e.g., an elementary school has no graduation data, so the averages for the indicators used only the assessment outcomes). District results are based on OSPI's aggregation rules, so the district results do not include results from correctional institutions, tribal schools, private schools or agencies

[5] Most of the other outcomes relate to high schools and the transition to higher education. Some data require transcript information, such as AP enrollment, dual enrollment, and college-ready rates. Other data sources could provide information about college entrance exams, college going rates, and remediation rates in higher education institutions.

[6] Very small schools (those with fewer than 10 assessed students) will have their index calculated but it will not be made public. However, the index will be viewed by state officials, and if the index is in the struggling tier on a consistent basis, the school could be placed in the Priority tier.

providing services, vocational schools/skill centers, schools that enroll more than 50% of their students from another district , and schools operated by a college or university that are not affiliated with a district.

## ACHIEVEMENT INDICATOR

This indicator has five outcomes: the four subjects tested by the WASL/WAAS statewide (reading, writing, math, and science) and the extended graduation rate (see explanation on how the rate is calculated below). The measure used is the percentage of "all" students meeting standard. Unlike the AYP measure, this indicator is what is shown on OSPI's online Report Card and does not reflect any adjustments (i.e., margin of error, continuous enrollment). The percent meeting standard includes both the results of the WASL and the WAAS, which is given to students with disabilities. These results are the combined total of the WASL and WAAS results found on the Report Card and are used when calculating AYP (without the margin of error and including students not continuously enrolled). For grade 10, only the first grade 10 attempt as reported in June of the tested year is used (this includes results for students who met standard in grade 9). Results from August assessments and retakes will be considered when looking at the "struggling" schools and districts to determine if they should be included in the Priority tier. This will recognize the districts that go to extra effort to help students who are in danger of not graduating unless they pass the required assessments. Subgroups results (for the various race/ethnicity groups, ELL, students with disabilities, gender) are used when examining the "struggling" schools and districts to determine if they should be included in the Priority tier. Results for low-income students are used in aggregate in a separate indicator described below.

The Achievement benchmarks and ratings for each of the four assessed subjects and the extended graduation rate are as follows:

- Achievement on *assessments* is scored based on the following percentage of students meeting standard:

  86-100% .....4
  70-85.9% .....3
  55-69.9% .....2
  40-54.9% .....1
  < 40% .....0

- Achievement on the *graduation rate* is scored based on the extended graduation rate from the previous year (see below for more information on how the graduation rate is calculated):

  > 95% .....4
  85-94.9% .....3
  75-84.9% .....2
  65-74.9% .....1
  < 65% .....0

Students from all tested grades in a school are combined for each subject, and the percentage of these students that meet standard on their respective tests is the school's percent meeting standard for that subject. This means the index can be calculated easily, regardless of a school's grade configuration (although grade configurations influence the results due to differences in the tests given). The same scoring benchmarks are used for all subjects. This gives equal importance to each subject.7 It also encourages the vertical alignment of the state assessments.

A school/district must have at least 10 students for it to be included in the accountability system. The minimum number used by OSPI is 10, but this policy is applied at the test and grade level. Using an N of 10 for a *school* means that very small schools will now be included in the accountability system because they will likely have at least 10 students assessed across the entire school. Combining all the test results together and using an N at the school level increases the overall N so a single student in a small school has less impact on the results and causes less of a change in the results from year to year. By using this system, scores in many schools that are currently suppressed at the grade level when there is less than 10 students assessed will become known in their aggregate form. This N policy means the state accountability system is more inclusive than the current AYP system, where the N is 30 and applies only students who are continuously enrolled. The advisors felt that the education system has a moral responsibility to serve all students, and having a small minimum N and counting students who have not been in class all year helps hold schools accountable for meeting the needs of *all* their students.

## ACHIEVEMENT VS. PEERS INDICATOR

This indicator uses the Learning Index (described below) level and controls for student characteristics beyond a school's control. Scores are the difference between the school's adjusted level and the average level among the school's peers. Specifically, the school/district score is the unstandardized residuals generated by a multiple regression. Those with scores above 0 are performing better than those with the same student characteristics, and those with scores below 0 are performing below those with the same student characteristics. The results are those for a single year rather than averages over multiple years for simplicity and to avoid the distortions when change takes place over time (e.g., when averaging, schools that have dramatic declines have better outcomes and schools with dramatic increases have worse outcomes).[8] Separate analyses were run for elementary, middle, high, and comprehensive (e.g., K-12) schools because of the variation of the

[7] The advisors did not have consensus about how to include science results in the index. Some felt that science should not be included at all because of changing standards and that it is not being taken seriously in many cases, which results in low scores across the state and relatively little improvement over time. As a result, it has little ability to differentiate school performance. Some suggested using lower cut points and raising them over time or including science but giving it less weight. After much discussion, a majority of the advisors concluded that since science will be a graduation requirement relatively soon, the only way to have science taken seriously was to treat it like the other subjects. Keeping the same rating system as the other subjects also keeps the system consistent and less complex and provides the opportunity to receive high ratings for improvement. Moreover, science achievement affects only two of the 20 cells of the matrix. Finally, not including science with equal weight penalizes those who work hard in this subject and sends the wrong message about the importance of students learning science concepts.

[8] In small schools, a single student could cause large changes in the index from year to year. However, analyses found relatively little difference in the amount of change in small schools compared to larger schools from one year to the next.

variables at each grade level. Schools serving specialized student populations (e.g., alternative schools, ELL and special education centers, private schools on contract, institutions) are not included in the regressions. Excluding these schools provides a better predicted level for the remaining regular schools in the analysis and better data for use when determining the cut scores for the various ratings. Since the specialized schools have such different characteristics, results for this indicator are not computed and their index is based on an average of their remaining ratings.

For schools, four student characteristics are the independent variables in the multiple regression: the percentage of (1) low-income students (percent eligible for free or reduced-price lunch[9]), (2) English language learners, (3) students with disabilities, and (4) mobile students (not continuously enrolled). A school's Learning Index from each of the four assessments (using WASL and WAAS results) as well as the extended graduation rate for high schools are the dependent variables. The regressions are weighted by the number of students assessed to prevent a small "outlier" school from distorting the regression (predicted) line. Although there is a high correlation between all the independent variables except special education, the regressions showed that all four variables helped improve the quality of the predicted levels, regardless of the regression method used.

For districts, three of the four student characteristics used in the school analysis were the independent variables in the multiple regression: the percentage of (1) low-income students (percent eligible for free or reduced-price lunch), (2) students with disabilities, and (3) mobile students (not continuously enrolled). The percentage of English language learners was not used because the initial analyses using this variable did not provide meaningful results. The same five dependent variables from the school-level analyses were used in the district analyses (the Learning Index for the four subjects and the extended graduation rate).

Financial information was also used as an independent variable in the district analysis. Funding data are available only at the district level, and some communities are able to raise higher levels of funding. The financial variable used is the total amount of operating revenue per weighted pupil. This variable controls for the level of funds available to the district. Weighting the student count "inflates" the enrollment figure because certain students require more resources to educate. The extra weight for ELL and low-income students was .20, which is the typical amount used in school finance studies (although the actual number is likely to be much higher). The weight for students with disabilities was .93, which is consistent with both the national research and the level of funding provided by the state.

- Achievement vs. Peers on the *assessments* is scored based on the difference between the actual and predicted Learning Index levels:

  > .20 .....4
  .10 to .20 .....3
  -.099 to .099 .....2
  -.20 to -.10 .....1
  < -.20 .....0

[9] The percentage of students in high schools who are eligible is often higher that what is reported, but this proxy for socioeconomic status is still the best available.

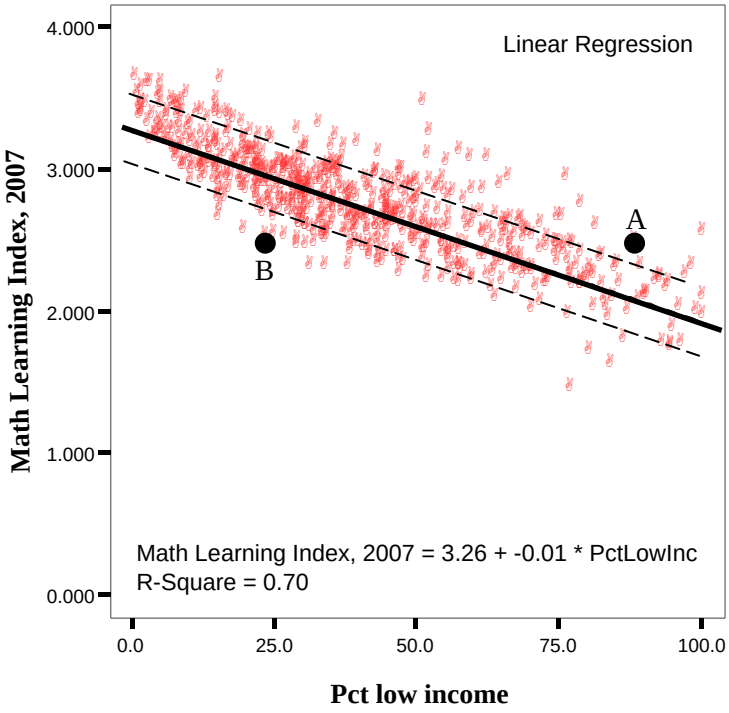
- Achievement vs. Peers on the *extended graduation rate* is scored based on the difference between the actual and predicted extended graduation rate:

  > 12 .....4
  5.01 to 12 .....3
  -5 to 5 .....2
  -5.01 to -12 .....1
  < -12 .....0

The mobility measure may need to be refined after further discussion takes place. Currently there is no common definition of mobility, and migrant student data does not include many students who are mobile. OSPI's student data system includes information about students who are/are not continuously enrolled from October 1 through the end of the testing period in May as part of the AYP system. Using this measure, the average state mobility rate is less than 6%. Most schools with mobility rates above 15% are alternative schools, and very few districts (mainly those in Pierce County close to military bases) have many of their schools with this high of a rate. However, the proposed measure may not identify students who move in and out of a school or district multiple times during the school year and are considered continuously enrolled, nor does it identify students that are new to the district and are still enrolled during the entire year. The proposed measure, the percentage of non-continuously enrolled students, can be used until a better measure is identified.

The scatterplot in Figure 9 illustrates how this indicator works. It shows just one of the independent variables (percent low income students) in relation to one outcome (K-6 math results). Each dot represents a school. The dark line is the average (predicted) level for a given Learning Index and low-income percentage. The distance between the school and the line is the difference from the predicted level. In this example, schools A and B have almost identical Learning Index results, but A falls well above the line while B falls well below the line. The dashed lines running parallel to the trend line represent the high and lowest cut points used for the ratings (.20 above and .20 below the trend line). When this kind of analysis is done factoring in the other variables (ELL, special education, mobility) at the same time in a multiple regression calculation, the distance from the predicted line is the school's score, which produces a rating. If the low-income variable was the only one used in the analysis, School A would have a rating of 4 because its index is more than .20 points above its predicted level, while school B would have a rating of 0 because its index falls more than .20 points below the predicted level.

**Figure 9: Scatterplot of Math Results in Elementary Schools by Percent Low Income**




The advisors discussed other possible independent variables that could be included in the analysis. These include the percentage of students who are enrolled in a gifted program, the percentage of minority students, and school size (enrollment).

- A gifted variable was not included because of a lack of reliable data, although the system should somehow take into account when a school has concentrations of these students. These schools will likely have very high index ratings.
- A race/ethnicity variable was not included because it is highly correlated with the other variables. Statistical analyses that included this variable found it added very little to the explanatory power of the model. Moreover, using this variable would reduce our ability to identify schools where students of color are treated differently. Finally, many of these students are also from low-income families, which is a separate indicator.
- A school size variable was not included because research findings to date reveal mixed results about how school enrollment levels affect student outcomes. School size is also a factor that can be controlled somewhat at the district level through the use of specialized programs and boundary lines. Other methods can be used to help schools compare themselves to those with similar sizes once the accountability results are made known.

The **Learning Index** is the dependent variable used for this indicator and for the Improvement indicator described below. This index, which was developed by the Commission on Student Learning and refined by the A+ Commission,[10] takes into consideration the percent of students performing at the different WASL levels. Specifically, the WASL and WAAS tests have five levels of performance:

[10] These Commissions are no longer in existence.

Level 0 – No score given[11]
Level 1 – Well below standard
Level 2 – Partially meets standard
Level 3 – Meets standard
Level 4 – Exceeds standard

This index is calculated like a grade point average with 4.0 as the highest score, reflects the level of student performance across the entire range of proficiency, not just those meeting standard. It gives greater weight to higher levels of proficiency on the state assessments and provides an incentive to support the learning of all students, including those well below standard (Level 1) and those that already meet the standard (Level 3) to they can move up to the next level. There is a "ceiling effect" when using this measure, but preliminary results show that even high-performing schools were achieving large gains because of the movement of students from Level 3 to Level 4. Once a school has all of its students in Level 4, there would not be any possibility to improvement any more, but all ratings together would still result in a school being in highest tier.

The following example shows how the Learning Index is calculated. The same method is used to calculate the index for all WASL tests (reading, mathematics, writing, science) in all the tested grades:

Level 0: 5% of all students assessed
Level 1: 15% of all students assessed
Level 2: 20% of all students assessed
Level 3: 40% of all students assessed
Level 4: 20% of all students assessed

**Learning Index = (0*0.05) + (1*0.15) + (2*0.20) + (3*0.40) + (4*0.20)**
**= 0 + .15 + .40 + 1.20 + .80 = 2.55**

## IMPROVEMENT INDICATOR

The Improvement indicator relies on changes in the Learning Index for the four assessed subjects and the graduation rate from one year to the next. Specifically:

- Improvement on *assessments* is scored on the levels of annual change in the Learning Index:

  > .12 .....4
  .051 to .12 .....3
  -.05 to .05 .....2
  -.051 to -.12.....1
  < -.12 .....0

- Improvement on *graduation rate* is scored on the level of percentage point change in the extended graduation rate from the previous year (see below for more information on how the graduation rate is calculated):

[11] The "No Score" designation includes unexcused absences, refusals to take the test, no test booklets but enrolled, incomplete tests, invalidations, and out-of-grade level tests.

> 6 .....4
3.01 to 6.00 .....3
-3.00 to 3.00 .....2
-6.00 to -3.01 .....1
< -6 .....0

A one-year change is used rather than using averages of previous years or a change from a year further in the past because it is the simplest calculation, it reflects the most recent set of results, and it does not distort the most recent results (using a two-year average helps a school if scores go down and penalizes the school if scores go up). New schools would only need two years of data to generate an improvement score. Since results are created each year, changes over time are seen when examining the results across multiple years.

The advisors discussed other possible improvement measures, including a 10% reduction in those not meeting standard (the AYP "safe harbor" measure), a 25% reduction in those not meeting standard over a 3-year period (the goal used for grade 4 reading several years ago), a percentage point gain from the previous year (or over several years), and a change in the scale score. While each of these have merit, the advisors determined that the annual change in the Learning Index provided the best measure of improvement because it focused on more than just those meeting standard and uses available data. The other measures can be used when analyzing "struggling" schools and districts for possible designation in the Priority tier.

## ACHIEVEMENT OF STUDENTS FROM LOW-INCOME FAMILIES

Much research has shown that student achievement is highly correlated with a family's socioeconomic status (SES). Specifically, academic achievement among students who live in low-income family is usually far below students from families that are not considered low income. This indicator focuses on the performance of low-income students. It uses the same five outcomes as the Achievement indicator: the four subjects tested by the WASL/WAAS statewide (reading, writing, math, and science) and the extended graduation rate. However, the outcome measures are the percentages of assessed students who are from low-income families who meet standard on the assessments and who graduate by the age of 21. The same rating scales are used as the achievement indicator.

Low-income status is measured in terms of the percentage of students who are eligible to receive a federally-subsidized meal (e.g., free or reduced-price lunch). The percentage of students in high schools who are eligible is often higher that what is reported, but this measure is still the best available proxy for SES. This indicator is highly correlated with the percentage of ELL students and students of color, two groups of students that often have lower levels of student achievement. The indicator is also positively correlated with students with disabilities and mobility.[12] The results for this indicator will not be different from the Achievement indicator if there are relatively few or no low-income students in a school.

## EXTENDED GRADUATION RATE MEASURE

[12] The statewide correlations between the percentage of students considered low-income and the percentage of students of color and ELL students in a school are .70 and .68 respectively. The correlations with mobility and special education are .49 and .27 respectively.

The Washington State definition of the on-time graduation rate is the percentage of students who graduate from public high school with a regular diploma (not including a GED or any other diploma not fully aligned with the state's academic content standards) in the standard number of years. The period of time required for students with disabilities to graduate is specified in each individualized education program (IEP). Students with disabilities who earn a diploma by completing the requirements of an IEP in the required period of time are counted as on-time graduates. The period of time required for EL and migrant students to graduate is determined on an individual basis when they enter the district and may be longer than the standard number of years. The period of time required to graduate for a migrant student who is not LEP and does not have an IEP can be one year beyond the standard number of years. LEP and migrant students who earn a diploma in the required period of time are counted as on-time graduates.

The on-time graduation rate is calculated as follows:[13]

*On-Time Graduation Rate* 100*(1-grade 9 dropout rate)*(1-grade 10 dropout rate)*(1-grade 11 dropout rate)*(1-grade 12 dropout rate-grade 12 continuing rate)

with *Dropout Rate* = $\frac{\text{number of students with a dropout, unknown, GED completer code}}{\text{total number of students served (less transfers out and juvenile detention)}}$

To encourage schools to serve students who remain in school beyond 4 years, a separate graduation rate is calculated that includes students who graduate in more than 4 years. This "**extended rate**" is be used for AYP purposes and the rate used in the accountability index. The formula for calculating this rate is as follows:

$$\textit{Extended Graduation Rate} = \frac{\text{number of on-time and late graduates}}{\text{\# of on-time graduates / on-time graduation rate}}$$

Dropouts are not counted as transfers. Since graduation data are not reported until after the beginning of the school year, the rates from the previous year are used.

The calculation method may change in the future when the state has enough data to track students over the entire time period. The cut scores for determining the ratings may need to change if another method produces substantially different results.

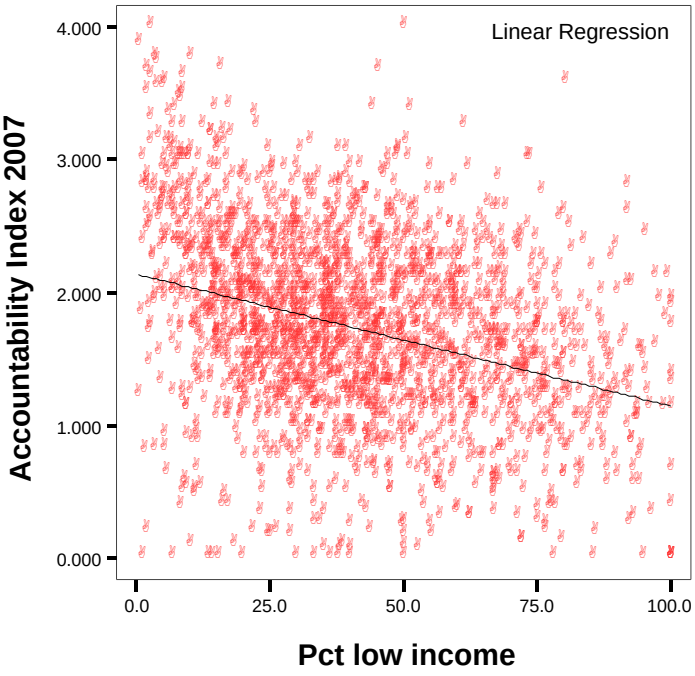
## DISTRIBUTION OF INDEX

Given the high correlation between family income and student performance, analyses were conducted to see how the school index related to the school's percentage of low-income students. Figure 10 shows these results for the 2,046 schools used in the analysis, while Figure 11 shows the results for the 296 districts. These figures show a much weaker relationship between the two variables than what would be seen if the dependent variable was

[13] See http://www.k12.wa.us/DataAdmin/pubdocs/GradDropout/03-04/Graduationanddropoutstatistics2003-04Final.pdf, chapter 1, for more information about these formulas.

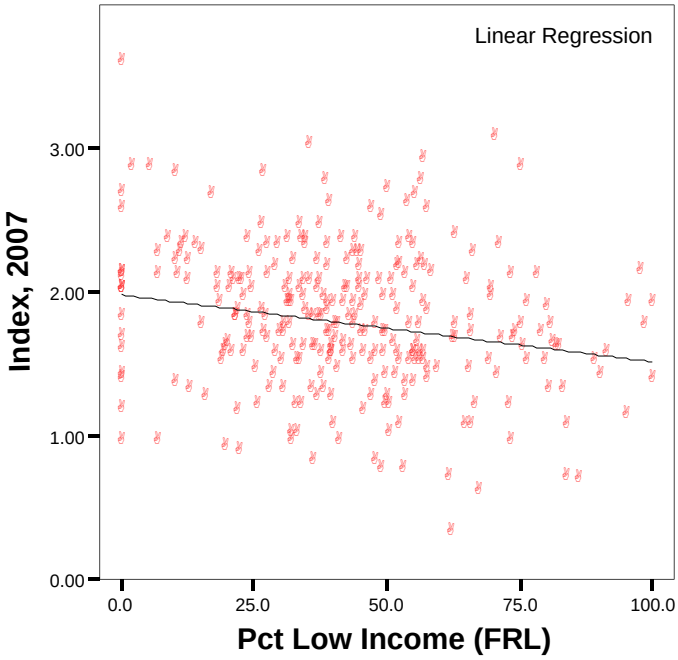
achievement. Many schools and districts that have relatively few low-income students still have rather low index scores, while many that have high concentrations of low-income students have rather high index scores. The trend line is still sloping downward, but the correlations and r-squares are relatively weak (-.33 and .11 for schools, -.22 and .05 for districts). These are much weaker than the relationship between student achievement and socioeconomic status. This is because achievement represents only half the index and is moderated by two of the other variables (improvement, peers) that have low correlations with socioeconomic status (all the school correlations with the improvement and peers variables were less than ± .08). It is harder for a school or district that has a high percentage of students who are low-income to achieve a very high index because the "all" students results are very similar to the low-income students results.

**Figure 10: Scatterplot of Index for All Rated Schools, by Percent Low Income**




**Accountability Index 2007 (rating average using student data) = 2.14 + -0.01 * PctLowInc**
**R-Square = 0.11**

**Figure 11: Scatterplot of Index for Districts, by Percent Low Income**




**Index, 2007 = 1.98 + -0.00 * Pct_LowIncome**
**R-Square = 0.05**

## RECOGNITION SYSTEM

Many of the guiding principles apply to the recognition system. The system should:

- Be transparent and simple to understand;
- Rely on multiple measures;
- Encourage the improvement of student learning and cooperation among educators;
- Focus at both the school and district levels;
- Rely mainly on criterion-referenced measures; and
- Provide multiple ways to demonstrate success and earn recognition.

With these principles in mind, the same matrix that is used to generate the index is also used to identify schools and districts for recognition. Cut points were developed for all 30 cells of the matrix after looking at distributions of the ratings for all schools. (The impact of the cut points on districts was not calculated for this analysis. District have fewer high ratings, as noted in Figures 1 and 5, so they would receive recognition less often than schools). To ensure recognition does not occur based on one good year alone, two years are averaged, and the average must meet minimum criteria.

Different cut points are used for different parts of the matrix because it is harder to achieve high ratings for some cells.

- For the “inner” 20 cells of the matrix, at least a 3.0 average is needed to receive recognition. To meet this level, a school/district needs to receive at least two straight ratings of 3, which are the second highest ratings (or it could have a rating of 2 & 4 in a

2-year period). Cells that average 3.5 or better (receive ratings of 3 & 4 or a 4 & 4) would receive recognition with "honors."

- For the 10 "averaged" cells on the outside of the matrix, at least a 2.75 is needed. This lower average is justified because it is much harder to achieve an average of 3.0 in the multiple categories. Relatively few schools and districts would be recognized at this lower level—on average only 14% of schools reached this level in each of the 10 cells, and even fewer districts reached this level (districts do not have as many high ratings). If a 3.0 were required instead of a 2.75, only about 9% of schools, on average, would receive recognition in these cells.
  - ✓ To meet an average of 2.75 in the *five outcome categories* (assessments and graduation rate), a school/district needs to have a total of 11 points in the four indicator ratings (11/4=2.75). This would usually require a majority of ratings of at least a 3 in two consecutive years.
  - ✓ To meet this level in the *four indicator categories* (achievement, improvement, achievement vs. peers, low-income achievement), a school/district needs to have a total of 14 points in the five outcome ratings (14/5=2.80). This would usually require 4 out of 5 ratings of at least a 3 in two consecutive years.
  - ✓ Like the "inner" cells of the matrix, any "averaged" cell with a 2-year average of 3.5 or better would receive recognition with "honors."

The number of schools and districts that receive recognition depends on the criteria described in Table 2. If the Board wanted to increase or decrease the amount of recognition provided, it could either change the criteria in Table 2 or change the cut points for recognition. Changes in the criteria in Table 2 would also affect the index scores for districts and schools. The Board could also request that a more formal "standard-setting" process take place to confirm or adjust the criteria used in Table 2.

The Board could establish additional criteria in order for a school/district to receive recognition. For example, the Board could require that recognition be given only if the achievement gap (e.g., between genders or between various groups of students) was decreasing. If could also require a closer analysis of the data before a school/district receives recognition with honors to ensure data problems (in their favor) or other factors are not responsible for very high ratings. This would prevent inappropriate designations that could undermine the accountability system.

A number of issues still need to be resolved related to the recognition. This includes what benefits accrue when a school or district meets the recognition criteria. The consequence could be as simple as highlighting the results on a Web site and issuing a press release about the winners. It could also generate financial rewards in certain cases. Another issue is what happens when a school and district are one in the same. The Board would need to make sure that any recognition is not duplicative (e.g., issuing a banner or financial reward for both the school and the district). Further, the proposed recognition should not duplicate existing awards being given by OSPI. Finally, the Board could create other types of recognition, such as special recognition for a few outstanding schools/districts and some that could be competitive in nature (e.g., require nominations or applications).

# APPENDIX B

## IDENTIFYING PRIORITY SCHOOLS AND DISTRICTS

The advisors (see Appendix D) helped generate a comprehensive list of quantitative and qualitative data that could be used to determine which schools in the "struggling" tier should be identified as needing more significant support from the state over a longer period of time (the Priority tier). Schools in the Priority tier would have the greatest need based on consistent underperformance on multiple measures (grades, subjects, indicators) over multiple years. The advisors assumed that being in this tier would generate the opportunity for substantially more support. The following factors were initially identified.

**Contextual Data**

- Type of school (alternative school, institution)
- Changes in student demographic profile (e.g., rapid increase in low-income or ELL students)
- What programs are included in the school (e.g., concentrations of ELL, special education, gifted)
- Program changes (e.g., establishing new ELL or special education programs)
- Student mobility
- Number of languages spoken by students
- Feeder schools
- Boundary changes (closures, consolidations)
- Construction or renovation projects

**Analysis of WASL/WAAS Results (annual and trends over time)**

- Achievement trends over multiple years for each subject area
- Size of the gap between WASL scores in different subjects
- Size of the achievement gap
- Percent students meeting 3 of 3 and 4 of 4 standards
- Trends for subgroups (gender, race/ethnicity, low-income) and programs (ELL, special education)
- Level of growth over time
- Changes in scale scores
- How performance compares to similar schools
- Results of students who have been in the school for longer periods of time (track cohorts of students to see how percent meeting standard changes over time, review results for just "continuously enrolled" students, the percentage of students meeting standard the next year in the next grade compared to the previous year, e.g., the percent in grade 4 in one year compared to the percent in grade 5 the next year)
- Results from retakes (high school) and collection of evidence

**AYP Results**

- Results generated with minimum Ns, confidence intervals, and continuously enrolled students (helps prevent false positives)
- How far the "all" group is from the annual goal
- Proficiency, participation, and other indicator results for all subgroups
- Number and percentage of cells not making AYP

- Which subgroups and subjects did not make AYP (ELL, special education, and participation rates count less, the all and race/ethnic groups count more)

**Other Quantitative Data (some may only be available at the district or school levels)**

- *Graduation data:* On-time and extended graduation rates for all students and subgroups, difference in rates, percentage of students still enrolled after four years
- *Dropout data:* Annual and cohort dropout rates for all students and subgroups, difference in rates
- *Discipline data:* Number of suspensions and expulsions, source of referrals, types of infractions, types of students being disciplined the most
- *Perception results:* Surveys of staff, parents, students about school conditions and how the results differ from one another
- *Classroom conditions:* Class sizes, student/teacher ratios by grade and subject
- *Staff characteristics:* Percentage of staff with certificates, teacher education/experience levels
- *Staff turnover:* Teacher and leadership changes at school and district levels
- *District assessments:* Results from any other assessments (e.g., MAP, grade 2 reading, portfolios)
- *WLPT results:* Performance of students from different language backgrounds, percentage of students exiting ELL program
- *Volunteers:* Number of parents volunteers, how they are used
- *Retention:* Number and percentage of students retained in grade, number and type of subjects not passed, level of credit deficiency
- *Finances:* Amount generated by local levies/bonds, fund balances, amount and sources of outside funding, stability in funding over time
- *District characteristics:* Number and percentage of schools in Tier 3, percentage of district students enrolled in Tier 3 schools
- *Data anomalies:* Incorrect data reported that could affect analyses, missing data, reason for missing data, number of ratings generating the average index

**Qualitative Data**

- *District role:* Resource amounts and types allocated to school, type of staff and programs provided, funding levels, type and intensity of interventions made to date, appropriateness of district policies, data analysis capacity, role of the district in school improvement efforts
- *Initiatives:* Number being attempted, focus and validity of initiatives, level of integration/cohesion among activities
- *Data use:* Quality of data system, capacity to use data, how information is used
- *Self-assessments:* Quality and use/implementation of school improvement plans
- *Staff relations:* Level of collaboration among staff and administrators within the school, union relations
- *Results from external reviews:* Results from accreditation and OSPI's Comprehensive Program Review (CPR), input from ESDs

Given the comprehensive nature of this list and the limited capacity to analyze all these data for every school and district in the "struggling" tier, the list was re-examined to determine which are the most important factors to analyze.

Schools serving special populations require separate analyses. For example, schools serving high concentrations of more challenging student populations (e.g., alternative schools,

institutions, those primarily serving ELL students and those with disabilities) often have low index results that would put them in the “struggling” tier. These schools have great need and should not be automatically excluded from being a Priority school. A closer look is needed to see if more support should be provided and the quality of programs serving these students. These kinds of schools may require an alternative accountability system (states like Texas have set up such a system). Some institutions should be excluded (e.g., jails & detention centers) but other included (e.g., long-term psychiatric facilities).

Other types of schools may need special analyses as well. For example, results for *very small schools* (N<10) are available but cannot be revealed to protect confidential information about students. However, the results could still be examined for trends over time. The number of *virtual schools* is increasing, often serving home-based students who are not required to take state assessments and may not be authorized to grant diplomas, which could mean there are few or no outcomes to measure. While some of these schools will generate results, they often serve many students outside the district, which means the school’s results are not included in the district results.

Certain preconditions need to exist for schools and district for them to use the additional resources effectively. For example, schools in the lowest tier need to be ready to benefit from the extra support. Without their buy-in, the chances for a successful reform are minimal. If the number of schools in the “struggling” tier is high and exceeds the level of resources available to support them, the state may want to consider using a minimum number of students per school to ensure cost-effectiveness of the assistance and allocating support by geographic location to ensure equity in distribution.

Finally, the schools and districts identified for the Priority tier may have a wide geographic distribution and be of different sizes. A single small school in a remote location may have the same level of need as a cluster of larger schools in a more accessible location. The state will need to determine how best to allocate its limited resources to ensure the cost effectiveness of its support.

# APPENDIX C

## CURRENT STATE ASSISTANCE PROGRAM

### SCHOOL IMPROVEMENT ASSISTANCE

The mission of the Office of Superintendent of Public Instruction's School Improvement Assistance (SIA) program is to help build capacity for districts and schools to improve student achievement through the use of the continuous school improvement model. This comprehensive model of support is unique in the United States. While many states have accountability systems that focus on rewards, punishments and takeovers, the SIA program provides comprehensive support for schools. Independent studies of the program have noted that the schools that received assistance for three years showed greater achievement gains than their respective comparison groups and the state as a whole. Nearly 60% of schools that have participated in SIA have exited federal improvement status and have made Adequate Yearly Progress (AYP) in the last two years of the program. The studies found further evidence that achievement gaps have been reduced in SIA schools.

Program Components

- **School Improvement Facilitator (SIF):** The facilitator works with OSPI, the school district, school, and a **School Improvement Leadership Team (SILT)** to develop a plan to address identified needs and to prepare and implement a jointly developed performance agreement between the school, school district and OSPI. The school improvement facilitators are experienced educators who have been successful in improving student performance and work approximately 1.5 days a week with each school for the three years of school improvement plan development and implementation. The school improvement leadership team includes representatives from the district and school staff, parents, and community members. Additional members may include educational service district (ESD) staff, OSPI staff and students.
- **Comprehensive Needs Assessment/School Performance Review:** The needs assessment/ school performance review is completed jointly by the school improvement leadership team, school district, OSPI, and a team of peer educators and experts. The school's strengths and challenges are identified and recommendations for improvement are developed. The school's curriculum, leadership, instructional practices and resources, assessment results, allocation of resources, parental involvement, support from the central office, and staff, parent, and student perceptions are examined. Student performance data, indicators from the "Nine Characteristics of High Performing Schools" and the results of a review of the school's reading and math instructional practices and program, are used to identify areas to consider for improvement. The assessment/audit includes the administration of survey instruments and an on-site visit.
- **School Improvement Process, Tools, and Support**: Schools are given the necessary processes, tools and expertise for the school improvement leadership team to develop a comprehensive *School Improvement Plan*. Funds are provided to contract with individuals to assist with components of the plan, and the school improvement facilitator are responsible for organizing and facilitating meetings in coordination with school and district staff.
- **Funds for Staff Planning and Collaboration:** Funds for planning time related to the development of the school improvement plan are provided. These funds may be used to

provide stipends for school improvement leadership team members. A minimum of three days must be devoted to planning time for all staff during the development of the school improvement plan. The funds can be used to pay staff stipends or to pay substitute teachers.

- **Performance Agreement**: Once the school improvement plan is completed, a two-year performance agreement is jointly developed by the school, school district and OSPI. The agreement identifies specific actions and resources the school district, the school and OSPI will commit to implement the school improvement plan. The agreement also includes a timeline for meeting implementation benchmarks and student improvement goals.
- **Implementation and Sustainability:** Tools and resources for the implementation of the performance agreement are provided during years two and three. The resources and expertise are determined on a case-by-case basis for each school, but could include such support as the provision of expertise in working with diverse student populations (e.g. special education, English language learners), funding and expertise to implement research-based practices and programs, and funding for time for staff collaboration. Schools and school districts are expected to ensure that existing funds are used effectively and to dedicate school district resources as identified in the jointly developed Performance Agreement.
- **Training Workshops**: Funds are provided to send a team of representatives to workshops during the school year to effectively plan for school improvement.
- **Professional Development:** Professional development opportunities for the school's principal and other school instructional leaders are provided in partnership with OSPI and the Association Washington School Principals (AWSP). Workshops are available during the school year.

The Process

***Year 1: School Improvement Planning and Performance Agreement***

- Conduct needs assessment through school performance review (formerly educational audit)
- Support staff training
- Develop school improvement plan/ performance agreement
- Develop student performance goals and evaluation criteria

***Year 2: Implementation***

- Tools and resources to implement the school improvement plan and performance agreements
- Evaluate student progress based on goals in the agreement

***Year 3: Sustainability***

- Tools and resources to build capacity and develop sustainability
- Evaluate student progress based on goals in the agreement

## DISTRICT IMPROVEMENT ASSISTANCE

For 2008-2009, districts fall in four district improvement groupings: (1) New in Step 1; (2) Continuing in Step 1; (3) New in Step 2; and (4) Continuing in Step 2. The technical assistance provided to districts in improvement status varies to meet the needs of districts either as they are developing their improvement plans or in various stages of implementation of their plans. The following areas are the most common types of support.

**A. Providing a School System Resource Guide (SSIRG):** OSPI and WASA collaborated in developing a resource planning guide that supports districts as they analyze existing

systems, structures, data, research findings, and more as they develop/revise their district improvement plan. A revision to the SSIRG is planned to be completed in 2008-09.

B. **Providing a Part-time, External District Improvement Facilitator:** District Improvement Facilitators are experienced educators who have been successful in improving student performance and receive continuous training through a partnership with WASA throughout the year. The selection of the facilitator is a collaborative effort between OSPI and each district. The facilitator works to help build the district's capacity to support high-quality, data-driven, research-based district improvement efforts.

C. **Providing or Arranging for Professional Development:** Additional resources for professional development to expand capacity of district and school personnel to sustain continuous improvement focused on improvement of instruction may be provided to meet the needs of districts.

D. **Provide for a District Educational On-Site Review:** Districts can request an educational on-site review to be completed by a team of peer educators and experts. The district's strengths and challenges are identified and recommendations for improvement are developed and provided to the district.

E. **Providing Identified Expertise:** Additional resources and expertise OSPI could provide is determined on a case-by-case basis for each district, but could include such support as expertise in working with diverse student populations (e.g., special education, English language learners), funding and expertise to implement research-based practices and programs, and funding for team collaboration time.

F. **Providing Limited Grant Money:** Districts may apply for two levels of grant support to assist in implementing one or more of the technical assistance opportunities listed A-E above.

OSPI recognizes the need to emphasize internal capacity building in districts and to revise its support systems and procedures over time.

# APPENDIX D

## ADVISORY GROUP MEMBERS

Dr. Pete Bylsma, an independent consultant and former state director of research and accountability at OSPI, was hired to help prepare the proposed index for Board review. He was assisted by a number of advisors. This diverse set of advisors reviewed the work that had been done to date, discussed numerous technical issues related to the proposed index, discussed the criteria for recognizing schools and districts, and identified quantitative and qualitative data that can be used to examine schools in the "struggling" tier to determine if they should be a Priority school needing much greater state assistance. Other stakeholders from OSPI were included in some of the discussions, and a State Board working group that focused on System Performance Accountability also provided feedback on the proposal.

Members of the advisory group were:

Ms. Maggie Bates, Hockinson SD (Assistant Superintendent)
Ms. JoLynn Berge, OSPI (Federal Policy and Grant Administrator)
Dr. Phil Dommes, North Thurston SD (Assessment Director)
Dr. Linda Elman, Tukwila SD (Assessment/Research Director)
Mr. Doug Goodlett, Vancouver SD (Special Services Director)
Dr. Peter Hendrickson, Everett SD (Assessment Director)
Dr. Feng-Yi Hung, Clover Park SD (Assessment/Evaluation Director)
Dr. Nancy Katims, Edmonds SD (Assessment Director)
Dr. Bill Keim, ESD 113 (Superintendent)
Ms. Linda Munson, South Kitsap SD (Special Programs Director)
Dr. Michael Power, Tacoma SD (Assistant Superintendent)
Mr. Bob Silverman, Puyallup SD (Executive Director for Assessment)
Ms. Nancy Skerritt, Tahoma SD (Assistant Superintendent)
Dr. Lorna Spear, Spokane SD (Executive Director for Teaching and Learning)
Dr. Alan Spicciati, Highline SD (Chief Accountability Officer)