

# *Executive Summary*

## State Accountability Index Report to the State Board of Education

---

May 21, 2009

Pete Bylsma, EdD, MPA  
BYLSMAPJ@COMCAST.NET

*The full report is available on the  
State Board of Education Web site at  
<http://www.sbe.wa.gov/spa.htm>.*

## CREATING THE ACCOUNTABILITY INDEX

The Legislature requires the State Board of Education (SBE) to develop a statewide accountability system to help improve academic performance among all students. SBE is required to “adopt objective, systematic criteria” 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 measures. The Board believes the index plays a key role in providing feedback about the status of education reform in schools and districts and in supporting continuous improvement efforts. Schools and districts in most need are given “Priority” status, making them eligible to receive more significant state support. These Priority schools and districts will be required to participate in a state system of support if initial offers of more support are not accepted and substantial improvement does not occur after several years. The creation of the index comes at a time when changes in the state’s assessment and data systems and at the U.S. Education Department provide an opportunity to consider new accountability ideas.

Various principles guided the development of the index. The index needs to (1) be transparent and simple to understand, (2) use existing data, (3) rely on multiple measures, (4) include assessment results from all grades and subjects tested statewide, (5) use concepts of the federal No Child Left Behind Act (NCLB) and its Adequate Yearly Progress (AYP) system when appropriate, (6) be fair, reasonable, and consistent, (7) be valid and accurate, (8) apply to both schools and districts, (9) apply to as many schools and districts as possible, (10) use familiar concepts when possible, (11) rely mainly on criterion-referenced measures instead of norm-referenced measures, (12) provide multiple ways to reward success, and (13) be flexible enough to accommodate future changes.

The proposed index is based on how schools and districts perform on a set 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: achievement of (1) non-low income students, (2) low-income students, (3) all students compared to those in similar schools/districts (controlling for the percentage of students who are learning English, have a disability, live in a low-income home, are mobile, and are designated as gifted), and (4) the level of improvement from the previous year. The results of the 20 measures form the 5x4 matrix shown in Table 1.

**Table 1: Matrix of Accountability Measures**

INDICATORS	OUTCOMES				
	Reading	Writing	Math	Science	Ext. Grad. Rate
Achievement of non-low income students					
Achievement of low income students					
Achievement vs. peers					
Improvement from the previous year					

Each cell of the matrix is rated on a 7-point scale (from 1 to 7) using fixed benchmarks. 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 students in all the

tested grades). **The index is the simple average of all the ratings** and ranges from 1.0 to 7.0. High schools and districts have 20 measures, while elementary and middle/junior high schools have only 16 measures because they do not have graduates. Table 2 shows how each of the five outcomes are measured using the four indicators and the benchmarks that produce the ratings.

**Table 2: Benchmarks and Ratings for Outcomes and Indicators**

	READING	WRITING	MATH	SCIENCE	EXT. GRAD. RATE <sup>1</sup>
<b>ACHIEVEMENT (NON-LOW INCOME)</b>	<u>% MET STANDARD</u>				<u>RATING</u>
	90 - 100% .....				7
	80 - 89.9% .....				6
	70 - 79.9% .....				5
<b>ACHIEVEMENT (LOW INCOME)</b>	60 - 69.9% .....				4
	50 - 59.9% .....				3
	40 - 49.9% .....				2
	< 40% .....				1
					<u>RATE</u>
					> 95..... 7
					90 - 95% ..... 6
					85 - 89.9% ..... 5
					80 - 84.9% ..... 4
					75 - 79.9% ..... 3
					70 - 74.9% ..... 2
					< 70%..... 1
<b>ACHIEVEMENT VS. PEERS<sup>2</sup></b>	<u>DIFFERENCE IN LEARNING INDEX</u>				<u>RATING</u>
	> .20.....				7
	.151 to .20.....				6
	.051 to .15 .....				5
	-.05 to .05 .....				4
	-.051 to -.15 .....				3
	-.151 to -.20 .....				2
	< -.20 .....				1
					<u>DIFFERENCE IN RATE</u>
					> 12..... 7
					6.1 to 12 ..... 6
					3.1 to 6..... 5
					-3 to 3..... 4
					-3.1 to -6 ..... 3
					-6.1 to -12 ..... 2
					< -12..... 1
<b>IMPROVEMENT<sup>3</sup></b>	<u>CHANGE IN LEARNING INDEX</u>				<u>RATING</u>
	> .15.....				7
	.101 to .15.....				6
	.051 to .10.....				5
	-.05 to .05 .....				4
	-.051 to -.10 .....				3
	-.101 to -.15 .....				2
	< -.15 .....				1
					<u>CHANGE IN RATE</u>
					> 6..... 7
					4.1 to 6..... 6
					2.1 to 4..... 5
					-2 to 2..... 4
					-2.1 to -4 ..... 3
					-4.1 to -6 ..... 2
					< -6..... 1

Note: Assessment results are the combined results from both the WASL and WAAS (assessments for students with disabilities) from all grades.

<sup>1</sup> This outcome only applies to schools and districts that are authorized to graduate students.

<sup>2</sup> This indicator adjusts the outcomes using statistical methods (multiple regression) to control for five student characteristics beyond a school’s control: the percentage of low-income, ELL, special education, gifted, and mobile students. (Mobile students are those who are not continuously enrolled from October 1 through the entire testing period.) Scores are the difference between the actual level and the predicted level of the Learning Index. Scores above 0 are “beating the odds” and negative scores are below the predicted level. Separate analyses are conducted for schools for each of the four assessments for each type of school (elementary, middle, high, multiple grade levels). District calculations also control for the level of current expenditures per pupil (adjusted for student need).

<sup>3</sup> Measured in terms of the change in the Learning Index from the previous year.

The proposed system holds *districts* accountable using the same indicators, outcomes, and criteria that are used for schools. The results are based on districtwide data for all grades rather than being disaggregated by grade bands (elementary, middle, high). In addition, financial data are used in the “peers” analysis to control for the amount of total operating expenditures per pupil (adjusted for student need).

**PRELIMINARY RESULTS**

Schools and districts fall into five tiers based on the index score. In-depth analyses of the data and conditions occurs for schools and districts that do not make AYP two years in a row to see if they merit further support. Those with the greatest need are eventually placed in a 6th (Priority) tier.

Table 3 shows the ranges for the tier assignments and the number of schools and districts that would have been placed in each tier in 2007 using the above criteria. Schools show a greater range than districts—far fewer districts were in the top and bottom tiers compared to the school results. The 228 schools in the Struggling tier enrolled 74,000 students (1 in 14 students statewide). Of the schools in this tier, 98 (43%) were alternative schools or served other special populations, and enrolled a total of 12,400 students. The 130 “regular” schools in the Struggling tier enrolled 61,600 students. Over the 2-year period, 149 schools (7.4%) had an average index below 2.50, and 89 were regular schools that enrolled approximately 39,000 students (roughly 4% of statewide enrollment). The 17 districts in the Struggling tier in 2007 tended to be rather small, averaging roughly 1,000 students. However, some larger districts had many schools in a struggling tier—17 districts had at least two regular schools and four districts had at least five regular schools with a 2-year index average below 2.50.

**Table 3: Tier Ranges and Preliminary Results (2007)**

Tier	Index Range	# of Schools	% of Schools	# of Students <sup>1</sup>	# of Districts	% of Districts	# of Students <sup>1</sup>
Exemplary	5.50 – 7.00	81	4.0%	28,650	1	.3%	360
Very Good	5.00 – 5.49	131	6.5%	64,500	9	3.1%	31,500
Good	4.00 – 4.99	591	29.4%	314,700	87	29.9%	278,500
Fair	2.50 – 3.99	980	48.7%	523,000	177	60.8%	692,500
Struggling	1.00 – 2.49	228	11.3%	74,000	17	5.8%	17,500
Priority <sup>2</sup>		TBD	TBD	TBD	TBD	TBD	TBD

<sup>1</sup>Approximate number (some schools did not provide enrollment data).

<sup>2</sup>To be determined after in-depth analyses of those not making AYP in at least two consecutive years.

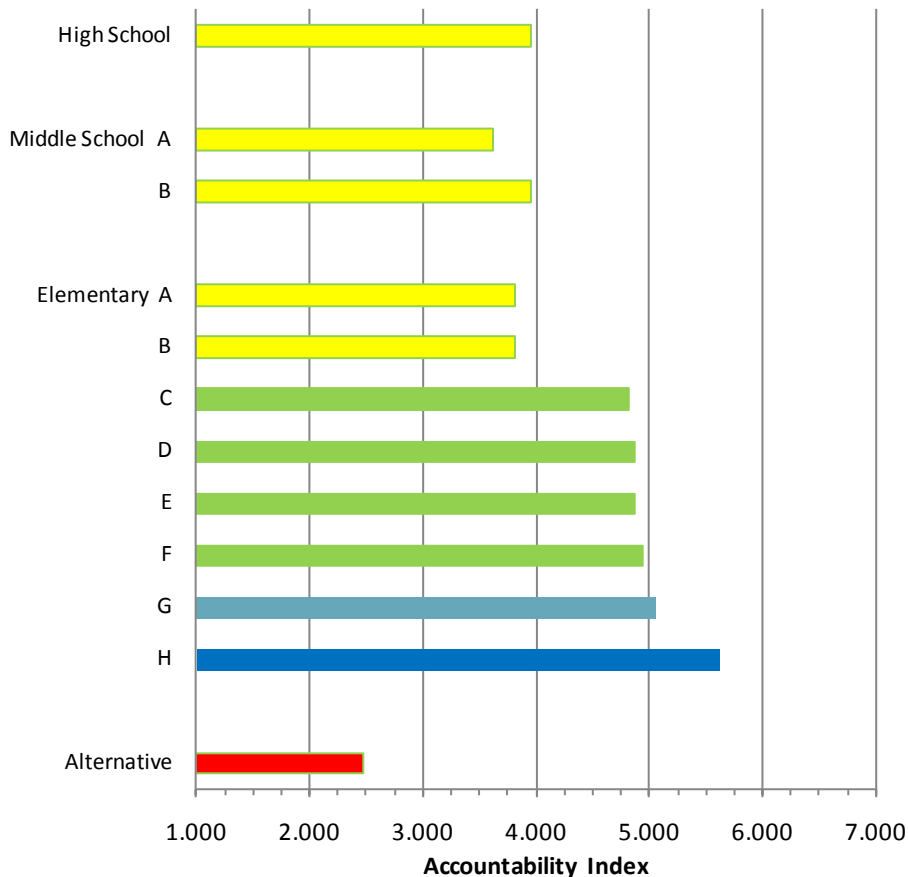
Table 4 provides an example of the ratings for an actual high school and how the average of the individual ratings generates the index/tier assignment. The school’s average rating of 3.40 is the index score, which puts the school in the middle of the Fair tier. The average ratings have been color-coded so the overall results can be seen at a glance. These types of results should be made public on the state Web site (the format for presenting the results must be determined). Results presented in this “dashboard” give policymakers, educators, and the public a quick snapshot of where a school is strong and weak, and it provides transparency about how the index is determined.

**Table 4: “Actual” High School, 2007**

Indicator	Reading	Writing	Math	Science	Grad Rate	Average
Non-low inc. ach.	5	6	3	1	5	4.00
Low-inc. ach.	4	4	1	1	7	3.40
Ach. vs. peers	2	2	2	2	6	2.80
Improvement	1	4	1	4	7	3.40
Average	3.00	4.00	1.75	2.00	6.25	3.40 ← INDEX

Various charts can illustrate district results as well. Figure 1 shows an example of how the index could be shown for each school in a district. In this example (an actual district), one school reached the Exemplary tier.

**Figure 1: Accountability Results in “Actual” District, 2008**



**HANDLING SPECIAL CASES**

The accountability system needs to be flexible enough to accommodate some special situations. These include holding alternative schools accountable using additional data, excluding some ELL results from the index calculations, and not counting the improvement cells when achievement is at very high levels.

Holding **alternative schools** accountable poses unique challenges. Many alternative schools exist in the state, and they vary greatly in their focus, structure, and clientele. Most are relatively small (total enrollment is less than 4% statewide), and more than half serve at-risk students in grades 9-12. Some believe these schools have taken on more challenging students while allowing traditional schools to generate better outcomes with their remaining students. On the other hand, some alternative schools offer special programs for students who are not at-risk and who meet rigorous academic requirements for admission. Some are considered “alternative” because they do not use a normal school approach, and a growing number of schools serve students through digital learning via the Internet. Parent Partnership Programs are a type of “school” where parents are the primary instructor and the district provides instructional support. Some target special student populations (e.g., special education, gifted, ELL). Given this variation, no “peer” indicator is computed for these schools.

Some alternative schools intentionally target student populations facing significant challenges and therefore are more likely to be in the Struggling tier. These schools should receive the normal index score based on calculations used for all schools. Alternative schools that do not make AYP in two consecutive years should be examined more closely to determine if they are using research-based best practices and showing progress. Areas for improvement should be identified and should be the focus of analysis if the alternative school does not make AYP again in the future.

**Results for ELL students** are currently included in AYP calculations in the student's second year of enrollment in a U.S. public school. OSPI has asked the U.S. Education Department to exclude ELL results until a student has been enrolled in a U.S. public school for three years or until the student achieves an advanced level of English proficiency on the WLPT, whichever comes first.<sup>1</sup> This request is based on research that shows it takes many years for an ELL student to acquire "academic" proficiency in English and because must be able to read and write English to understand and respond to each test item. Moreover, testing students who do not understand English violates widely-adopted testing standards because of threats to validity and mistreatment of human subjects. However, the Department has denied OSPI's repeated request to use this policy.

Nevertheless, computing the accountability index should exclude the results for ELL students who have not achieved advanced proficiency (Level 3 composite) on the WLPT or who are in their first three years of enrolling in a U.S. public school, whichever comes first, for any test that requires reading and writing in only English.<sup>2</sup> In addition, OSPI should begin reporting WLPT results on its Report Card in a way that allows educators, parents, and other stakeholders to monitor the progress ELLs make in terms of learning English and meeting state content standards. Finally, all ELLs should be required to take the WASL after their first year of enrollment, and OSPI should analyze the WASL and WLPT results to determine the extent to which ELLs are on track to meet state standards.

Most ELL results would still be included in the accountability index, even with this "extended exclusion, because (1) most ELLs enter school in kindergarten and have attended school for three years before taking state assessments for the first time in grade 3, and (2) most ELLs enrolled in the assessed grades (3-8 and 10) reach the advanced level of the WLPT. As a result, the exclusion has little impact on the index results. Nevertheless, the combination of recommendations improves the validity of the accountability system and provides more information about the progress of ELLs.

**Schools and districts that perform at very high levels** are not able to improve much from the previous year. To avoid "penalizing" these schools for a lack of improvement, the ratings for this indicator should not be included in the index calculations under certain conditions. Without this policy, schools/districts with nearly all of their students achieving Level 4 on an assessment and graduating nearly all their students would not be able to achieve a rating above 4 (little or no improvement). Specifically, the improvement indicator should be excluded when computing the index whenever a Learning Index reaches 3.85 out of 4.00 and remains at or above that level for two consecutive years. (A school or district needs to improve by more than .15 to receive a rating of 7, which is impossible when their Learning Index reaches 3.85.) The first year the Learning Index falls above 3.85, a school/district would get a rating based on its improvement. If the Index stays at or

---

<sup>1</sup> The composite score from the annual Washington Language Proficiency Test (WLPT) reflects proficiency in reading, writing, speaking, and listening. The three-year exemption period reflects the views of most stakeholders and is the average time required for ELL students to meet standard on the WASL.

<sup>2</sup> The math and science tests are available in Spanish and Russian for the first time in 2009 but responses must be made in English.

above 3.85, the maximum rating is not possible and the indicator should not be calculated.<sup>3</sup> The same policy applies to the extended graduation rate outcome (when the rate reaches or exceeds 94% in two consecutive years.<sup>4</sup>

## INTEGRATING THE FEDERAL AND STATE ACCOUNTABILITY SYSTEMS

Federal law requires states to have a single accountability system. Many states have combined their state accountability system with the federal NCLB system. However, stakeholders across Washington believe the federal system is overly complex and that the AYP results do not provide an accurate picture of school and district quality. As a result, stakeholders are working to develop a set of policies that would use the index to determine AYP and different consequences for schools and districts that do not make AYP over an extended period of time. When these policies are finalized in the coming months, SBE and OSPI will submit a unified accountability plan to the U.S. Education Department that will recommend using the state accountability system for federal accountability purposes. A new administration may provide more flexibility to states that design alternative systems that provide more rigorous and valid accountability. All the data current reporting requirements of NCLB would continue to be met (i.e., making public the disaggregated data for the assessments, participation, and “other indicators” for the various student subgroups). Moreover, new data elements would be made public to further increase the rigor of the system.

### Advantages Over the Current System

The proposed accountability system has many desirable features that make it a preferred alternative to the current rules used to measure AYP while simultaneously increasing the system’s rigor.

- The index is a *more valid* measure of school and district performance because it is based on the performance of all students in more subjects, is more nuanced than a Yes/No (pass/fail) system, and addresses several unintended consequences created by the current AYP system.
- The index is *more inclusive/comprehensive* because it uses a smaller minimum number for reporting (10 students across the entire school/district), includes the results of all students (not just those continuously enrolled through the testing period), includes both writing and science (this helps prevent a narrow curriculum), and uses the Learning Index to measure performance across the range of assessment results (reduces the focus on “bubble” students who perform close to the proficiency cut point at the expense of students who are farther above and below that level).
- The index is *less volatile* over time because assessment results are combined across all grades in a school and district rather than using results for individual grades where students change from one year to the next.
- The index is *more transparent* because it does not include a margin of error, the benchmarks are the same over time and among the different subjects for both schools and districts, there are fewer subgroups and rules, and schools and districts have the same minimum number required for reporting the results.
- Using the index to determine AYP helps the state *maintain high performance standards*. Two recent studies found that Washington has some of the nation’s toughest AYP requirements, resulting in a high percentage of schools not making AYP and undergoing sanctions.<sup>5</sup> Using the

---

<sup>3</sup> Of the schools and districts with reportable data (at least 10 students assessed), two schools reached this level on the Learning Index in 2006 (one in both reading and writing and the other in writing), but no district reached this level in 2006 and no school or district reached this level in any subject in 2007.

<sup>4</sup> Of the schools with graduation data, 11% had a rate that was at least 94% in two consecutive years.

<sup>5</sup> See “*The Accountability Illusion*,” Thomas B. Fordham Institute (February 2009) and “*Schools Struggling to Meet Key Goal on Accountability*,” Education Week (January 7, 2009).

index reduces the incentive for the state to lower its standards so all students can be counted as proficient and meet federal targets, which are viewed as unrealistic if standards are kept high.

### Identifying Schools and Districts Needing Improvement

Each fall OSPI will compute the accountability index and apply the rules for making AYP. All schools and districts in all tiers will be given an AYP status, not just those receiving Title I funds. The first time a school or district does not make AYP, it is in a “warning” year. Schools and districts that do not make AYP two years in a row should not automatically fall into “improvement” status. Instead, they should undergo an in-depth review by OSPI staff (this is different from the current system where federal rules dictate an automatic designation). Professional judgment panels can be used to conduct this review. The results of this review would determine if the school/district should move into an “improvement” step and be required to take certain actions.

The data to be reviewed fall in five general categories. The list below provides examples in each.

- **Contextual Data**
  - Type of school
  - Changes in student population
  - Programs served by the school
- **Assessment Results (WASL/WAAS/WLPT)**
  - Trends over multiple years for each subject area
  - Subgroup trends (e.g., race/ethnicity, ELL, special education)
  - Results for students who have been enrolled for at least two years
- **Federal AYP Results**
  - Participation rates for all subgroups
  - “Other indicator” data (unexcused absence and graduation rates) for all subgroups
- **Teaching and Learning Issues**
  - Teacher education and experience levels
  - Student/teacher ratio
  - Recent changes in leadership (key central office staff and principals) and teachers
  - Alignment of curriculum and materials across grades and with state standards
- **Other Data**
  - Graduation and dropout rates for subgroups
  - Funding from local levies/bonds and outside sources
  - Problems with data that generate the index (e.g., reporting errors related to graduates)

### Consequences and the Priority Tier

NCLB currently requires schools and districts to undergo increasing levels of “sanctions” if they do not make AYP over an extended period of time. NCLB also requires schools and districts that are in an improvement step to make AYP two years in a row in order to exit improvement status. Many stakeholders believe the sanctions and exit criteria are flawed and need to be changed. A different set of consequences will be proposed after consulting with OSPI and stakeholders statewide.

While the Accountability Index cannot be used to determine AYP in 2009, it can still be calculated and made public so the details of the index can be used for educational purposes and by OSPI in its assistance decisions. Eventually, schools and districts with the most significant need should be placed in the Priority tier and offered significant state support that is tailored to meet their specific

needs (participation is voluntary). If extra assistance is not accepted and improvement does not occur, a binding corrective action plan would be established between the district and the state, if authorized by the Legislature.<sup>6</sup>

## **RECOGNITION**

Index results can be calculated retroactively and used for recognition purposes. Providing recognition in Fall 2009 would be considered “Phase I” in the implementation of the accountability system, with full implementation contingent upon the provision of adequate funding. Using the index in this way will provide a more valid picture of school/district performance than AYP results, and it will introduce the concepts to the various stakeholders prior to its full implementation.

The recognition system should (1) be transparent and simple to understand, (2) rely mainly on criterion-referenced measures, and (3) provide multiple ways to demonstrate success. The recognition system is based on the belief that people are motivated more by success than by blame or guilt, and they need clear, challenging, and attainable goals.

SBE and OSPI are working together to create a unified recognition system based on the index results. At least two forms of recognition should be considered. The first is for “Outstanding Overall Performance” and the second is for “Noteworthy Performance.” The form of recognition given should depend on the difficulty of reaching the award criteria: recognition for Outstanding Overall Performance should have a high profile, while recognition for Noteworthy Performance should be handled in an inexpensive and efficient manner.

### **Outstanding Overall Performance (8 types)**

SBE should provide recognition based on high levels of performance in eight areas: the index, each of the five outcome areas, for closing the achievement gaps (a minimal difference between non-low income and low income ratings in all subjects), and for a small percentage of schools with high levels of gifted students. To ensure only truly outstanding performance is recognized, schools and districts should meet the following conditions.

- (a) For the **index**, the 2-year average must be at least 5.50 and there must be fewer than 10% students designated as gifted each year.
- (b) For **reading, writing, math, science**, and the **extended graduation rate**, the overall (column) 2-year average must be at least 6.00, at least 2 of the 4 cells in the column must be rated each year, and there must be fewer than 10% students designated as gifted each year.
- (c) For the **achievement gap**, there must be at least 10 students in at least 2 of the 5 outcomes (columns) in both of the income-related cells (non-low income and low income), there can be no rating of 1 in any income-related cell or peer cell, there can be no more than a 1-point difference in the rating between the two income-related cells,<sup>7</sup> and there must be fewer than 10% students designated as gifted each year. Each of the above criteria must be met two years in a row.
- (d) For **schools with gifted programs**, the top 5% of schools in grade band—elementary, middle, high, and multi-level—that have at least 10% gifted students would receive this type of recognition, based on the 2-year average peer ratings in all four subjects.

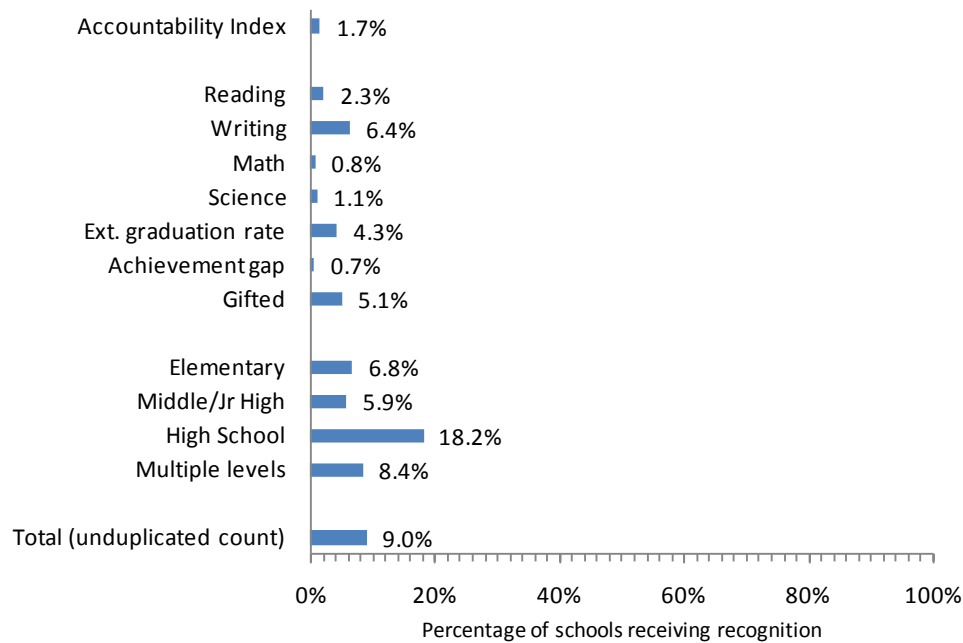
---

<sup>6</sup> ESHB 2261, passed by the 2009 Legislature, contains language on this issue.

<sup>7</sup>For example, if the reading non-low income cell is rated 5, the reading low-income cell could be rated no lower than 4 and no higher than 6.

Figure 1 shows the percentage of all schools that met the criteria in 2008. If the system were in place, recognition would have been given to 191 different schools in a total of 277 areas (some schools would have received recognition in more than one area). This represents 9% of all schools. This level of recognition is similar to the OSPI’s School of Distinctions award.

**Figure 1: Percentage of Schools Meeting “Outstanding Overall Performance” Criteria (2008)**



Noteworthy Performance (21 types)

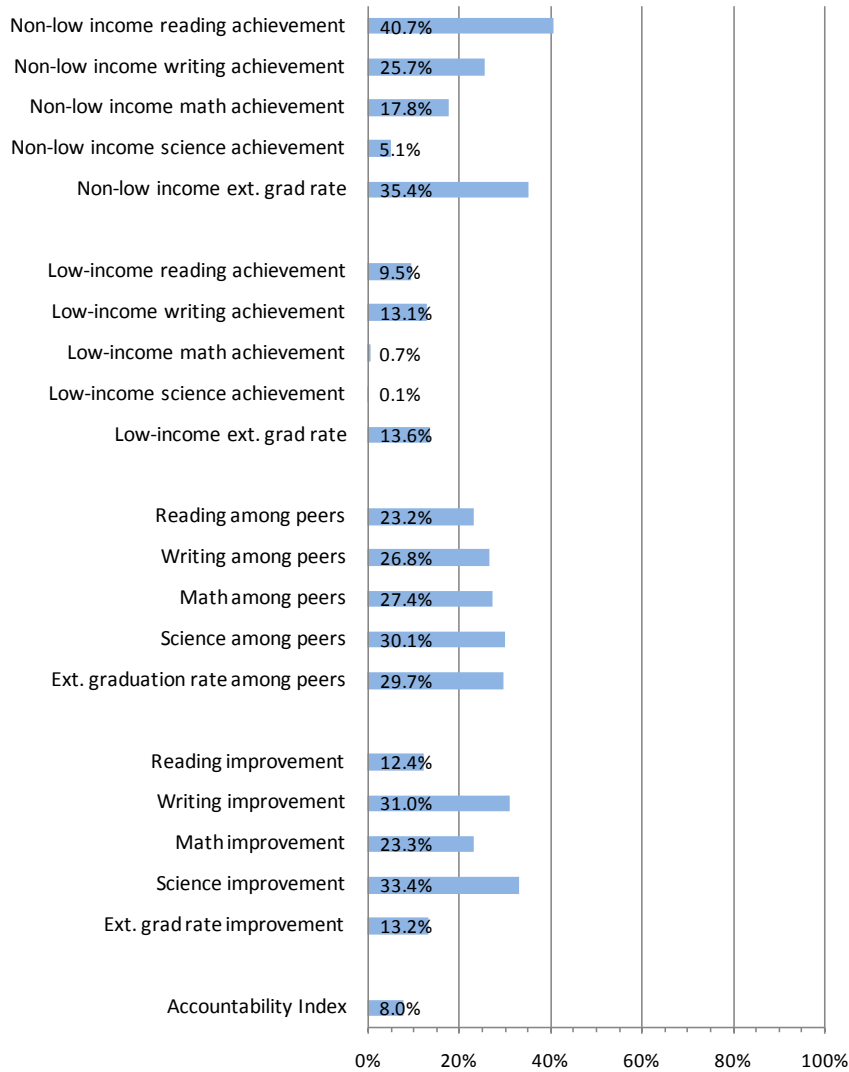
OSPI should consider giving recognition to schools and districts for each of the 20 cells of the matrix when the 2-year average for a cell is at least 5.50, and for the index when the 2-year average is at least 5.00. To receive this type of recognition, schools and districts should also meet the following conditions.

- (a) No rating below 5 can occur in either year in the 20 cells of the accountability matrix.
- (b) Recognition for non-low income cells in reading and writing requires a minimum 2-year average of the low income group of 4.00.

This option provides recognition to far more schools because it is based on performance in each of the 20 cells of the matrix as well as the index. More than 80% of the schools statewide (1,618 in total) met the criteria in some way in 2008, and some schools would have received recognition for performance in many of the cells of the matrix.

Figure 2 shows the percentage of schools that met the criteria for recognition in the 21 cells in 2008. Some areas would have received more recognition than others. The largest number of schools (40%) met the minimum criteria for non-low income reading achievement (even when requiring the low income group to have at least a 4.0 average). Achievement in math, science, and among low-income students had far fewer schools meeting the criteria. For the index, 8% had an overall 2-year average of at least 5.00.

**Figure 2: Percentage of Schools Meeting “Noteworthy Performance” Criteria (2008)**



Other forms of recognition could be given by OSPI or SBE based on their priorities. For example, OSPI could recognize a certain percentage of schools in math and science, even if they do not meet the criteria discussed above. Monetary compensation is not recommended, although matrix data could be used to generate schoolwide bonuses if the Legislature includes these as part of any law or reforms of the basic education finance system in the future.

**REMAINING WORK**

A number of issues must still be resolved before the index can be implemented effectively. Various OSPI and SBE activities need to be integrated and aligned with one another (e.g., how the index relates to NCLB requirements, how to use the index to identify Priority schools and districts, how and when assistance and recognition occur, how index results are represented and made available to the public). The methods for holding alternative schools accountable need further development. Finally, the proposed accountability system will need to remain flexible in order to adapt to changes in NCLB and graduation requirements, the assessment system and content standards, and other factors that may impact the results.